PCT WORLD INTELLECTUAL PROPEI
International Bu

INTERNATIONAL APPLICATION PUBLISHED UNDER

(51) International Patent Classification 6 :
A61K 47/48, 41/00, C07K 14/475, 19/00, C12N 1/21, 15/12

A2

(11) Int

WO 9608274A2

(43) International Publication Date: 21 March 1996 (21.03.96)

(21) International Application Number: PCT/US95/12205

(22) International Filing Date: 13 September 1995 (13.09.95)

(30) Priority Data:

08/305,771	13 September 1994 (13.09.94)	US
08/441,979	16 May 1995 (16.05.95)	US

(81) Designated States: AM, AU, BB, BG, BR, BY, CA, CN, CZ, EE, FI, GE, HU, IS, JP, KG, KP, KR, KZ, LK, LR, LT, LV, MD, MG, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TT, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).

Published
Without international search report and to be republished upon receipt of that report.

(71) Applicant: PRIZM PHARMACEUTICALS, INC. [US/US]; 11035 Roselle Street, San Diego, CA 92121-1204 (US).

(72) Inventors: SOSNOWSKI, Barbara, A.; 1013 Adella Avenue, Coronado, CA 92118 (US). CHANDLER, Lois, Ann; 1814 Blue Bonnet Place, Encinitas, CA 92014 (US). HOUSTON, L., L.; 4045 Carmel View #92, Del Mar, CA 92014 (US). NOVA, Michael, P.; P.O. Box 3023, Rancho Santa Fe, CA 92067 (US). McDONALD, John, R.; 4440 Mistral Place, San Diego, CA 92130 (US).

(74) Agents: NOTTENBURG, Carol et al.; Seed and Berry, 6300 Columbia Center, 701 Fifth Avenue, Seattle, WA 98104-7092 (US).

(54) Title: CONJUGATES OF HEPARIN-BINDING EPIDERMAL GROWTH FACTOR-LIKE GROWTH FACTOR WITH TARGETED AGENTS

(57) Abstract

Conjugates of heparin-binding epidermal growth factor-like growth factor (HBEGF) linked, either directly or via a linker, to a targeted agent are provided. The targeted agent is a cytotoxic agent, such as a ribosome-inactivating protein (RIP) and an antisense nucleic acid, or is a therapeutic nucleic acid for targeted delivery. The targeted agent is attached to HBEGF, or via a linker, through a chemical bond, or the conjugate is prepared as a chimera using techniques of recombinant DNA. The conjugates are used to target cytotoxic agents or therapeutic nucleotides to cells bearing EGF receptors and are particularly useful for treating solid tumors, such as breast and bladder tumors, and for treating disorders involving pathophysiological proliferation of smooth muscle cells, such as restenosis.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

Description

CONJUGATES OF HEPARIN-BINDING EPIDERMAL GROWTH FACTOR-LIKE GROWTH FACTOR WITH TARGETED AGENTS

Technical Field

This invention is related to the preparation and use of heparin-binding epidermal growth factor-like growth factor (HBEGF) conjugated to a targeted agent, such as a cytotoxic protein or a nucleic acid. The conjugates are for use as anti-tumor agents, for the treatment of disorders involving pathophysiological proliferation of smooth muscle cells, such as restenosis, and to effect genetic therapy of cells that bear receptors for heparin-binding epidermal growth factor.

Background of the Invention

A major goal of treatment of neoplastic diseases and hyperproliferative disorders is to ablate the abnormally growing cells while leaving normal cells untouched. Various methods are under development for providing treatment, but none provide the requisite degree of specificity.

One method of treatment is to provide toxins. Immunotoxins and cytotoxins are protein conjugates of toxin molecules with either antibodies or factors which bind to receptors on target cells. Three major problems may limit the usefulness of immunotoxins. First, the antibodies may react with more than one cell surface molecule, thereby effecting delivery to multiple cell types, possibly including normal cells. Second, even if the antibody is specific, the antibody reactive molecule may be present on normal cells. Third, the toxin molecule may be toxic to cells prior to delivery and internalization. Cytotoxins suffer from similar disadvantages of specificity and toxicity. Another limitation in the therapeutic use of immunotoxins and cytotoxins is the relatively low ratio of therapeutic to toxic dosage. Additionally, it may be difficult to direct sufficient concentrations of the toxin into the cytoplasm and intracellular compartments in which the agent can exert its desired activity.

Given these limitations, cytotoxic therapy has been attempted using viral vectors to deliver DNA encoding the toxins into cells. If eukaryotic viruses are used, such as the retroviruses currently in use, they may recombine with host DNA to produce infectious virus. Moreover, because retroviral vectors are often inactivated by the complement system, use *in vivo* is limited. Retroviral vectors also lack specificity in delivery; receptors for most viral vectors are present on a large fraction, if not all, cells. Thus, infection with such a viral vector will infect normal as well as abnormal cells.

Because of this general infection mechanism, it is not desirable for the viral vector to directly encode a cytotoxic molecule.

While delivery of nucleic acids offers advantages over delivery of cytotoxic proteins such as reduced toxicity prior to internalization, there is a need for high specificity of delivery, which is currently unavailable with the present systems.

In view of the problems associated with gene therapy, there is a compelling need for improved treatments which are more effective and are not associated with such disadvantages. The present invention exploits the use of conjugates which have increased specificity and deliver higher amounts of nucleic acids to targeted cells, while providing other related advantages.

Summary of the Invention

The present invention generally provides conjugates of heparin-binding epidermal-like growth factor-like growth factor (HBEGF) polypeptide or a portion thereof and a targeted agent. In one embodiment of this invention, the HBEGF and targeted agent are conjugated through a linker. Within each conjugate, there can be more than one HBEGF and targeted agent molecule. Preferably, in the conjugates, there are between one and six HBEGF and targeted agents, and most preferably one HBEGF molecule and one targeted agent. In certain embodiments, the linker is selected from the group consisting of protease substrates, linkers that increase the flexibility of the conjugate, linkers that increase the solubility of the conjugate, photocleavable linkers and acid cleavable linkers. In certain other embodiments, the HBEGF polypeptide may be mammalian HBEGF or HBEGF that is modified by addition of a cysteine residue or replacement of a nonessential amino acid residue within about 20 amino acids of the N-terminus or C-terminus. In yet other embodiments, the targeted agent is cytotoxic, preferably a ribosome inactivating protein, and most preferably saporin. Other cytotoxic agents include a nucleic acid.

In another embodiment, the conjugate has the formula: targeted $agent_n$-$(L)_q$-$HBEGF_m$ or $HBEGF_m$-$(L)_q$ targeted $agent_n$, wherein n and m, which may be the same or different, are at least 1.

In another aspect, methods of treating HBEGF-mediated pathophysiological conditions, comprising administering to the animal a therapeutically effective amount of a conjugate between HBEGF and a cytotoxic agent, are provided. In certain embodiments, the condition is a dermatological disorder involving epidermal cells, a neoplastic disorder of epidermal or mesodermal cells, an ophthalmic disorder involving proliferation of epithelial cells, or a disorder characterized by proliferation of

smooth muscle cells. Methods are also provided to inhibit proliferation of cells bearing HBEGF receptors, comprising contacting the cells with an effective amount of a HBEGF targeted agent conjugate.

In yet other aspects, methods of effecting gene therapy are provided, wherein cells are contacted with a conjugate having a targeted agent which is a nucleic acid, and the conjugate includes a nuclear translocation sequence linked to the targeted nucleic acid or HBEGF.

In yet other aspects, DNA fragments, encoding a conjugate between a targeted agent and HBEGF are provided. In certain embodiments, the DNA conjugate may additionally comprise a linker. Plasmids, vectors, and host cells are also provided. In another embodiment, methods of producing a fusion protein of HBEGF and a targeted agent are provided comprising (a) culturing cells transformed with a plasmid containing a DNA fragment according to claim 21, under conditions whereby the DNA fragment is transcribed and translated; (b) lysing the cells in a buffer containing urea; (c) eluting the protein from a cation-exchange chromatography resin; (d) passing the protein over an anion-exchange chromatography resin; (e) eluting the protein from a cation-exchange chromatography resin; (f) eluting the protein from a hydrophobic interaction chromatography resin; and (g) recovering the protein from a size exclusion chromatography resin.

In other embodiments, the HBEGF is modified by insertion of a cysteine residue within about 20 amino acids of the N-terminus or C-terminus, wherein the inserted residue replaces a nonessential residue in the unmodified HBEGF.

Pharmaceutical compositions, comprising the HBEGF targeted agent conjugate and a physiological acceptable excipient are also provided.

These and other aspects of the present invention will become evident upon reference to the following detailed description and attached drawings. In addition, various references are set forth below which describe in more detail certain procedures or compositions, and are therefore incorporated by reference in their entirety.

Detailed Description of the Invention

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which the subject matter herein belongs. All U.S. patents and all publications mentioned herein are incorporated in their entirety by reference thereto.

The "amino acids" are identified according to their well-known, three-letter or one-letter abbreviations. The nucleotides, which occur in the various DNA fragments, are designated with the standard single-letter designations used routinely used in the art.

As used herein, to "bind to a receptor" refers to the ability of HBEGF to detectably bind to such receptors, as assayed by standard *in vitro* assays. For example, binding, as used herein, measures the capacity of the HBEGF conjugate or HBEGF polypeptide to specifically bind to a HBEGF receptor (known as EGF receptor) on smooth muscle or epidermoid cells, such as A431 cells, using a procedure substantially as described in Moscatelli (1987) *J. Cell Physiol. 131*:123-130. Briefly, cells are grown to subconfluence and incubated in appropriate buffer with detectably labeled, such as radioiodinated HBEGF polypeptide in the presence of various concentrations of an HBEGF polypeptide of interest. Binding affinity is measured by counting the membrane fraction that is solubilized in a suitable buffer containing a detergent, such as in 0.5% Triton X-100 in PBS (pH 8.1).

As used herein, the term "biologically active," or reference to the "biological activity of a cytotoxic conjugate of HBEGF," such as a conjugate containing HBEGF and saporin refers to the ability of such polypeptide to enzymatically inhibit protein synthesis by inactivation of ribosomes either *in vivo* or *in vitro* or to inhibit the growth of or kill cells upon internalization of the saporin-containing polypeptide by the cells. Such biological or cytotoxic activity may be assayed by any method known to those of skill in the art including, but not limited to, the *in vitro* assays that measure protein synthesis and *in vivo* assays that assess cytotoxicity by measuring the effect of a test compound on cell proliferation or on protein synthesis. Particularly preferred, however, are assays that assess cytotoxicity in targeted cells.

As used herein, "biological activity" refers to the *in vivo* activities of a compound or physiological responses that result upon *in vivo* administration of a compound, composition or other mixture. Biological activity, thus, encompasses therapeutic effects and pharmaceutical activity of such compounds, compositions and mixtures. Such biological activity, however, may be defined with reference to particular *in vitro* activities, as measured in a defined assay. Thus, for example, reference herein to the biological activity of HBEGF or fragment thereof refers to the ability of the HBEGF to bind to cells bearing HBEGF receptors and internalize a linked agent. Such activity is typically assessed *in vitro* by linking the HBEGF (or fragment) to a cytotoxic agent, such as saporin, contacting cells bearing HBEGF receptors, such as A431 cells, with the conjugate and assessing cell proliferation or growth. Such *in vitro*

activity should be extrapolatable to *in vivo* activity. *In vivo* activity may be assessed using recognized animal models, such as the mouse xenograft model for anti-tumor activity (*see, e.g.*, Beitz et al. (1992) *Cancer Research 52*:227-230; Houghton et al. (1982) *Cancer Res. 42*:535-539; Bogden et al. (1981) *Cancer (Philadelphia) 48*:10-20; Hoogenhout et al. (1983) *Int. J. Radiat. Oncol., Biol. Phys. 9*:871-879; Stastny et al. (1993) *Cancer Res. 53*:5740-5744).

As used herein, a "conjugate" refers to a molecule that contains at least one HBEGF moiety and at least one targeted agent that are linked directly or via a linker and that are produced by chemical coupling methods or by recombinant expression of chimeric DNA molecules to produce fusion proteins.

As used herein, the term "cytotoxic agent" refers to a molecule capable of inhibiting cell function. The agent may inhibit cell growth, differentiation or proliferation or may be toxic to cells. This term includes agents that, when internalized by a cell, interfere with or detrimentally alter cellular metabolism or in any manner inhibit cell growth or proliferation. The term includes agents whose toxic effects are mediated when transported into the cell and also those whose toxic effects are mediated at the cell surface. A variety of cytotoxic agents can be used and include those that inhibit protein synthesis and those that inhibit expression of certain genes essential for cellular growth or survival.

As used herein, "DNA encoding an HBEGF peptide or polypeptide" refers to any DNA fragment encoding an HBEGF, as defined above. Exemplary DNA fragments include: any such DNA fragments known to those of skill in the art; any DNA fragment that encodes an HBEGF that binds to an HBEGF receptor and is internalized thereby and may be isolated from a human cell library using any of the preceding DNA fragments as a probe; and any DNA fragment that encodes any of the HBEGF polypeptides set forth in SEQ ID NOs. 2-5. Such DNA sequences encoding HBEGF fragments are available from publicly accessible databases, such as: DNA* July, 1993 release from DNASTAR, Inc. Madison, WI, and Genbank Accession Nos. M93012 (monkey) and M60278 (human); the plasmid pMTN-HBEGF (ATCC #40900) and pAX-HBEGF (ATCC #40899) described in published International Application WO/92/06705 (see, also, the corresponding U.S. Patent upon its issuance); and Abraham et al. (1993) *Biochem. Biophys. Res. Comm. 190*:125-133). DNA encoding HBEGF polypeptides will, unless modified by replacement of degenerate codons, hybridize under conditions of at least low stringency to DNA encoding a native HBEGF (*e.g.*, SEQ ID NO. 1). In addition, any DNA fragment that may be produced from any of the preceding DNA fragments by substitution of degenerate codons is also

sequence of HBEGF polypeptides, and DNA fragments encoding such peptides, are available to those of skill in this art, it is routine to substitute degenerate codons and produce any of the possible DNA fragments that encode such HBEGF polypeptides. It is also generally possible to synthesize DNA encoding such peptides based on the amino acid sequence.

As used herein, a "fusion protein" refers to a polypeptide that contains at least two components, such as a HBEGF polypeptide and a targeted agent, a HBEGF polypeptide and linker, or a HBEGF polypeptide, linker, and targeted agent, and is produced by expression of DNA in a host cell.

As used herein, "heparin-binding epidermal growth factor-like growth factor" (HBEGF) polypeptides refer to any polypeptide that binds to an HBEGF receptor, and is transported into the cell by virtue of its interaction with the receptor. Native HBEGF has a heparin-binding domain. In particular, HBEGF refers to polypeptides having amino acid sequences of a native HBEGF polypeptide, as well as HBEGF polypeptides modified by amino acid substitutions, deletions, insertions or additions in the native protein, but retains the ability to bind to a HBEGF receptor and to be internalized in a cell bearing such receptor. Such HBEGF polypeptides include, but are not limited to human HBEGF (SEQ ID NO. 2), monkey HBEGF (SEQ ID NO. 4) and rat HBEGF (SEQ ID NO. 5). Reference to HBEGFs is intended to encompass HBEGF polypeptides isolated from natural sources as well as those made synthetically, as by recombinant means or by chemical synthesis. This term also encompasses the precursor forms, such as those set forth in SEQ ID NOs. 1, 2, 4, and 5, and mature forms, such as that set forth in SEQ ID No. 3. HBEGF also encompasses muteins of HBEGF that possess the ability to target a targeted agent, such as a cytotoxic agent, including but not limited to ribosome- inactivating proteins, such as saporin, light activated porphyrin, and antisense nucleic acids, to HBEGF-receptor expressing cells. Muteins of HBEGF include, but are not limited to, those produced by replacing one or more of the cysteines with serine as described herein or those that have any other amino acids deleted or replaced, as long as the resulting protein has the ability to bind to HBEGF-receptor bearing cells and internalize the linked targeted agent. Typically, muteins will have conservative amino acid changes, such as those set forth below in Table 1. DNA encoding such muteins will, unless modified by replacement of degenerate codons, hybridize under conditions of at least low stringency (1 X SSPE or SSC, 0-.1% SDS, 50°C, medium stringency; 0.2 X SSPE or SSC, 0.1% SDS, 50°C; high stringency; 0.1 X SSPE or SSC, 0.1% SDS, 65°C) to DNA encoding native HBEGF (*e.g.*, SEQ ID NO. 1) and encode an HBEGF polypeptide, as defined herein.

As used herein, "mature HBEGF" refers to processed HBEGFs. Various isoforms of mature HBEGF have variable N-termini, and include, but are not limited to, those having N-termini corresponding to amino acid positions 63, 73, 74, 77 and 82 of the precursor protein (*see, e.g.,* SEQ ID Nos. 1, 2, *see also* SEQ ID Nos. 4 and 5). As used herein a "portion of a HBEGF" refers to a fragment or piece of HBEGF that is sufficient to bind to a receptor to which native HBEGF binds and internalize a linked targeted agent.

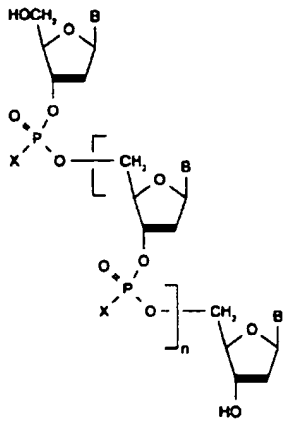
As used herein, "HBEGF-mediated pathophysiological condition" refers to a deleterious condition characterized by or caused by proliferation of cells that are sensitive to HBEGF mitogenic stimulation. HBEGF-mediated pathophysiological conditions include, but are not limited to, conditions involving pathophysiological proliferation of smooth muscle cells, such as restenosis, certain tumors, such as solid tumors including breast and bladder tumors, tumors involving pathophysiological expression of EGF receptors, dermatological disorders, such as psoriasis, and ophthalmic disorders involving epithelial cells, such as recurrence of pterygii and secondary lens clouding.

As used herein, the "HBEGF receptor" (HBEGF-R) refers to a receptor that reacts with members of the HBEGF family of proteins and that is able to transport HBEGF into the cell. For example, HBEGF polypeptides interact with the high affinity EGF receptors (EGF-R) on bovine aortic smooth muscle cells and A431 epidermoid carcinoma cells (see Higashiyama et al. (1991) *Science 251*:936-939; Higashiyama et al. (1992) *J. Biol. Chem. 267*:6205-6212). Thus, EGF-receptors, which are also HBEGF-Rs, include the EGF receptors described in U.S. Patent Nos. 5,183,884 and 5,218,090, Ullrich et al. (1984) *Nature 309*:418-425, those encoded by the *erbB* gene family.

As used herein, "nucleic acids" refer to RNA or DNA that are intended as targeted agents, which include, but are not limited to, DNA encoding therapeutic proteins, fragments of DNA for co-suppression, DNA encoding cytotoxic proteins, antisense nucleic acids and other such molecules. Reference to nucleic acids includes duplex DNA, single-stranded DNA, RNA in any form, including triplex, duplex or single-stranded RNA, anti-sense RNA, polynucleotides, oligonucleotides, single nucleotides and derivatives thereof.

Nucleic acids may be composed of the well-known deoxyribonucleotides or ribonucleotides composed of the bases: adenosine, cytosine, guanine, thymidine, and uridine. As well, various other nucleotide derivatives and non-phosphate backbones or phosphate derivative backbones may be used.

For example, because normal phosphodiester oligonucleotides (referred to as PO oligonucleotides or type I; see structure, below, where X = 0) are sensitive to DNA- and RNA-specific nucleases, several resistant types of oligonucleotides have been developed (*see, e.g.*, International Application WO 93/23570, which is based on 07/881,255, filed May 11, 1992; International Application WO 93/15742, which is based on 07/833,146, filed February 10, 1992; Wagner et al. (1993) *Science 260*:1510-1514; U.S. Patent No. 5,218,088, U.S. Patent No. 5,175,269; U.S. Patent No. 5,109,124; Carter et al. (1993) *Br. J. Cancer 67*:869-876); these include types II-IV:



in which B is a nucleotide base; and X is OEt in phosphotriester (type II), X is Me in methylphosphonate (type III; referred to as MP oligos); and X is S in phosphorothioate (referred to as PS oligos; U.S. Patent No. 5,218,088 to Gorenstein et al. describes a method for preparation of PS oligos). Presently, MP and PS oligonucleotides have been the focus of most investigation.

As used herein, a "therapeutic nucleic acid" describes any nucleic acids used in the context of invention that modify gene transcription or translation. This term also includes nucleic acids that bind to sites on proteins and to receptors. It includes, but is not limited to the following types of nucleic acids: nucleic acids encoding a protein, antisense RNA, DNA intended to form triplex molecules, extracellular protein binding oligo nucleotides and small nucleotide molecules. A therapeutic nucleic acid may serve as a replacement for a defective gene or encode a therapeutic product, such as TNF or a cytotoxic molecule, such as saporin. The therapeutic nucleic acid may encode all or a portion of a gene, and may function by recombining with DNA already present in a cell, thereby replacing a defective portion of a gene. It may also encode a portion of a protein and exert its effect by virtue of co-suppression of a gene product.

As used herein, "restenosis" refers to a process and the resulting condition that occurs following angioplasty in which the arteries become reclogged. After treatment of arteries by balloon catheter or other such device, denudation of the interior wall of the vessel occurs, including removal of the endothelial cells that constitute the lining of the blood vessels. Smooth muscle cells (SMCs), which form the blood vessel structure, proliferate and fill the interior of the blood vessel. This process and the resulting condition is restenosis.

As used herein, "substantially homologous" with reference to an HBEGF polypeptide means that the protein is more homologous (*i.e.*, shares more amino acid residues in common) to any of the mature HBEGF polypeptides included in SEQ ID Nos. 1-6 than is TGF-α. With reference to DNA it means that the DNA encodes a substantially homologous protein, and, but for substitution of degenerate codons, hybridizes under conditions of at least low stringency to DNA encoding any of the mature HBEGFs included in the sequences set forth in SEQ ID Nos. 1-6.

As used herein, isolated, "substantially pure DNA" refers to DNA fragments purified according to standard techniques employed by those skilled in the art (*see, e.g.*, Maniatis et al. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY and Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.).

As used herein, a "targeted agent" is any agent that is intended for internalization by linkage to a targeting moiety, as defined herein, and that upon internalization in some manner alters or affects cellular metabolism, growth, activity, viability or other property or characteristic of the cell. The targeted agents include proteins, polypeptides, organic molecules, drugs, nucleic acids and other such molecules. As used herein, to target a targeted agent means to direct it to a cell that expresses a selected receptor by linking the agent to a polypeptide reactive with a HBEGF receptor. Upon binding to the receptor the targeted agent or targeted agent linked to the HBEGF is internalized by the cell.

A. Heparin binding epidermal growth factors

1. Polypeptides reactive with an HBEGF receptor

For the purposes of this invention, HBEGF need only bind a specific HBEGF receptor and be internalized. Any member of the HBEGF family, whether or not it binds heparin, is useful within the context of this invention as long as it meets the requirements set forth above. Members of the HBEGF family are those that have sufficient nucleotide identity to hybridize under normal stringency conditions (typically greater than 75% nucleotide identity). Subfragments or subportions of a full-length HBEGF may also be desirable. One skilled in the art may find from the teachings provided within that certain biological activities are more or less desirable, depending upon the application.

Heparin-binding EGF-like growth factors (HBEGFs) are mitogens in the epidermal growth factor (EGF) protein family that bind to the glycosaminoglycan,

heparin. HBEGFs elute from heparin-Sepharose columns at about 1.0 - 1.2 M NaCl and were first identified as a secreted product of cultured human monocytes, macrophages, and the macrophage-like U-937 cell line (*see, e.g.*, Higashiyama et al. (1991) *Science 251:*936-939; Besner et al. (1990) *Cell Regul. 1:*811-819). HBEGFs, also called "heparin-binding EGF-homologous mitogen (HB-EHM)" (see WO 92/06705), are a family of growth factors that have a broad spectrum of activities, including a mitogenic effect on a variety of cells, such as BALB/c 3T3 fibroblast cells and smooth muscle cells. HBEGFs, however, are not mitogenic for endothelial cells (Higashiyama et al. (1991) *Science 251:*936-939).

As isolated, mature HBEGFs are heterogeneous in structure and contain up to 86 amino acids, including two sites of *O*-linked glycosylation (Higashiyama et al. (1992) *J. Biol. Chem. 267:*6205-6212). The carboxyl-terminal half of the secreted human HBEGF shares approximately 35% sequence identity with human EGF, and includes six cysteines spaced in the pattern characteristic of members of the EGF protein family. HBEGF interacts with the same high affinity receptors as EGF on bovine aortic smooth muscle cells and human A431 epidermoid carcinoma cells (*see, e.g.*, Higashiyama (1991) *Science 251:*936-939). The amino-terminal portion of the mature factor, which includes stretches of hydrophilic residues, has no structural equivalent in EGF. The heparin-binding residues of HBEGF reside primarily in a twenty one-amino acid stretch upstream of and slightly overlapping the EGF-like domain. HBEGF appears to be a more potent mitogen for smooth muscle cells than either EGF or TGF-α, which also binds to EGF receptors.

Mammalian HBEGFs are derived from a 208 amino acid precursor protein. The human and monkey precursor proteins share 97% sequence identity, the rat and mouse precursors are 92% identical; and there is 80% sequence identity between primate and rodent HBEGF precursor proteins (see Abraham et al. (1993) *Biochem. and Biophys. Res. Comm. 190:*125-133). The mature HBEGF polypeptides are heterogenous and range from about 75-86 amino acids in length. HBEGFs have a molecular weight of approximately 19-23 kD, and have an isoelectric point between about 7.2-7.8.

The effects of HBEGFs are mediated at least in part by receptor tyrosine kinases on the cell surface membranes of HBEGF-responsive cells (*see, e.g.*, U.S. Patent Nos. 5,183,884 and 5,218,090; and Ullrich et al. (1984) *Nature 309:*418-425, which are incorporated herein by reference). The EGF receptor proteins, which are single chain polypeptides with molecular weights of approximately 170 kD, depending on cell type, constitute a family of structurally related EGF receptors. Cells that express

the EGF receptors include, for example, smooth muscle cells, fibroblasts, keratinocytes, and numerous human cancer cell lines, such as the: A431 (epidermoid); KB3-1 (epidermoid); COLO 205 (colon); CRL 1739 (gastric); HEP G2 (hepatoma); LNCAP (prostate); MCF-7 (breast); MDA-MB-468 (breast); NCI 417D (lung); MG63 (osteosarcoma); U-251 (glioblastoma); D-54MB (glioma); and SW-13 (adrenal). HBEGFs also bind to the heparan sulfate proteoglycans, which appear to internalize bound moieties via the endocytic pathway and contribute to internalization of HBEGFs.

For purposes herein, polypeptides that are reactive with a HBEGF receptor include any molecule that (1) includes a receptor binding domain that is homologous to EGF and that is substantially homologous (more homologous than TGF-α) to such domains in the mature HBEGFs having amino acid sequences set forth in SEQ ID Nos. 1-5; and (2) reacts with receptors on cells to which a native HBEGF (a mature HBEGF having an amino acid sequence included in any of SEQ ID Nos. 1-5) and results in internalization of the linked agent. Thus, the polypeptides that are reactive with a HBEGF receptor include members of the HBEGF family of polypeptides, muteins of these polypeptides, chimeric or hybrid molecules that contain portions of any of these family members, and any portion thereof that binds to HBEGF receptors and internalizes a linked agent. Any polypeptide that has a heparin-binding domain and includes an EGF receptor binding domain that is substantially homologous (more homologous than TGF-α) to such domains set forth in any of SEQ ID Nos. 1-5 is intended for use herein. HBEGF for use herein also includes any fragment of a HBEGF polypeptide that retains the ability to bind to a HBEGF receptor and to be internalized by a cell bearing such receptors.

It is understood that minor amino acid sequence variations including allelic variations, species variations and conservative amino acid substitutions, such as those set forth in Table 1, in HBEGF that do not alter its ability to bind to HBEGF receptors and to be internalized by cells upon such binding are encompassed within the family of HBEGF polypeptides intended for use herein.

Mature human HBEGF as isolated has heterogenous amino acid lengths ranging from 75-86 (Higashiyama et al. (1992) *Science 251*:936-939). For example, various isoforms of mature HBEGF that have variable N-termini, and include, but are not limited to, those having N-termini corresponding to amino acid positions 63, 73, 74, 77 and 82 of the precursor protein (*see, e.g.*, SEQ ID Nos. 1 and 2, see, also SEQ ID No. 3, for the presently preferred form). A preferred HBEGF for use herein is the 77 amino acid form of human HBEGF beginning at amino acid 73 of the precursor protein (SEQ ID No. 3, which corresponds to amino acids 73-149 of SEQ ID NOs. 1 and 2; see

Example 4). Members of the HBEGF polypeptide family, including SEQ ID NOs. 2-5, are particularly preferred. Modification of the polypeptide may be effected by any means known to those of skill in this art. The preferred methods herein rely on modification of DNA encoding the polypeptide and expression of the modified DNA.

All of the HBEGF polypeptides induce mitogenic activity in a wide variety of cells, and this activity is mediated by binding to an HBEGF cell surface receptor followed by internalization. Binding to a HBEGF receptor followed by internalization are the activities required for an HBEGF polypeptide to be suitable for use herein; mitogenic activity is not required. A test for binding and internalization activity is the ability of the HBEGF-toxin conjugates to kill EGF-receptor containing cells. An exemplary method for testing for such cytopathic activity is the Cell Proliferation/Cytotoxicity Assay described in Example 4. Any HBEGF polypeptide that possesses such ability is intended for use herein.

2. Modifications of HBEGF

If it is necessary or desired, the heterogeneity of preparations of HBEGF polypeptide-containing chemical conjugates can be reduced by modifying the HBEGF polypeptide by deleting or replacing a site(s) (that are non-essential for binding and internalization) on the HBEGF that cause the heterogeneity and/or by modifying the targeted agent. Such sites are typically cysteine residues that, upon folding of the protein, remain available for interaction with other cysteines or for interaction with more than one cytotoxic molecule per molecule of HBEGF polypeptide, but that are not required for binding to HBEGF receptors and internalization. Such cysteine residues do not include any cysteine residue that are required for proper folding of the HBEGF polypeptide, or for retention of the ability to bind to a HBEGF receptor and internalize. For chemical conjugation, one cysteine residue that, in physiological conditions, is available for interaction, is not replaced because it is used as the site for linking the cytotoxic moiety. The resulting modified HBEGF is conjugated with a single species of targeted agent, such as a RIP, antisense nucleic acid or therapeutic nucleic acid.

The contribution of each cysteine to the ability to bind to HBEGF receptors may be determined empirically. Each cysteine residue may be systematically replaced with a conservative amino acid change or deleted. The resulting mutein is tested for the requisite biological activity, the ability to bind to HBEGF receptors and internalize linked targeted moieties. If the mutein retains this activity, then the cysteine residue is not required. Additional cysteines are systematically deleted and replaced and the resulting muteins are tested for activity. In this manner the minimum number

and identity of the cysteines needed to retain the ability to bind to a HBEGF receptor and internalize may be determined.

The HBEGF polypeptide may also be modified by addition of one or more cysteine residues at or near the C- or N-terminus, preferably the N-terminus, in order to render it more amenable to chemical conjugation by providing a readily available non-essential cysteine residue. HBEGF is modified herein by addition of Cys residues at or near the N-terminus in order to render them more amenable for chemical conjugation. Any HBEGF may be modified for use herein by replacement of one or more cysteine residues that are not required for binding to a HBEGF receptor and internalization of the targeted agent. These modified forms of HBEGF are particularly suitable for chemical conjugation to linkers and/or targeted agents.

Mutation may be effected by any method known to those of skill in the art, including site-specific or site-directed mutagenesis of DNA encoding the protein and the use of DNA amplification methods using primers to introduce and amplify alterations in the DNA template, such as nucleic acid amplification splicing by overlap extension (SOE). Site-specific mutagenesis is typically effected using a phage vector that has single- and double-stranded forms, such as M13 phage vectors, which are well-known and commercially available. Other suitable vectors that contain a single-stranded phage origin of replication may be used (*see, e.g.*, Veira et al. (1987) Meth. Enzymol. 15:3). In general, site-directed mutagenesis is performed by preparing a single-stranded vector that encodes the protein of interest (*i.e.*, a member of the HBEGF family or a cytotoxic molecule, such as a saporin). An oligonucleotide primer that contains the desired mutation within a region of homology to the DNA in the single-stranded vector is annealed to the vector followed by addition of a DNA polymerase, such as *E. coli* polymerase I Klenow fragment, which uses the double stranded region as a primer to produce a heteroduplex in which one strand encodes the altered sequence and the other the original sequence. The heteroduplex is introduced into appropriate bacterial cells and clones that include the desired mutation are selected. The resulting altered DNA molecules may be expressed recombinantly in appropriate host cells to produce the modified protein.

The SOE method uses two amplified oligonucleotide products, which have complementary ends as primers and which include an altered codon at the locus at which the mutation is desired, to produce a hybrid product. A second amplification reaction that uses two primers that anneal at the non-overlapping ends amplify the hybrid to produce DNA that has the desired alteration.

Suitable conservative substitutions of amino acids are known to those of skill in this art and may be made generally without altering the biological activity of the resulting molecule. Those of skill in this art recognize that, in general, single amino acid substitutions in non-essential regions of a polypeptide do not substantially alter biological activity (*see, e.g.*, Watson et al. *Molecular Biology of the Gene*, 4th Edition, 1987, The Benjamin/Cummings Pub. co., p.224). Such substitutions are preferably made in accordance with those set forth in TABLE 1 as follows:

TABLE 1

Original residue	Conservative substitution
Ala (A)	Gly; Ser
Arg (R)	Lys
Asn (N)	Gln; His
Cys (C)	Ser; neutral amino acids
Gln (Q)	Asn
Glu (E)	Asp
Gly (G)	Ala; Pro
His (H)	Asn; Gln
Ile (I)	Leu; Val
Leu (L)	Ile; Val
Lys (K)	Arg; Glny; Glu
Met (M)	Leu; Tyr; Ile
Phe (F)	Met; Leu; Tyr
Ser (S)	Thr
Thr (T)	Ser
Trp (W)	Tyr
Tyr (Y)	Trp; Phe
Val (V)	Ile; Leu

Other substitutions are also permissible and may be determined empirically or in accord with known conservative substitutions. Any such modification of the polypeptide may be effected by any means known to those of skill in this art.

HBEGF polypeptides may be isolated by methods known to those of skill in the art or may be prepared by expression of DNA encoding a HBEGF

polypeptide (*see, e.g.*, International Application WO/92/06705 (and the corresponding U.S. patent application serial No. 07/598,082), and Abraham et al. (1993) *Biochem. Biophy. Res. Comm. 190*:125-133 and SEQ ID NOs. 1-5 herein).

B. Targeted agents

1. Cytotoxic agents

Cytotoxic agent refers to a molecule capable of inhibiting cell function. Cytotoxic agents include any agent that, upon internalization, by a eukaryotic cell, inhibits growth or proliferation of the cell, either by killing the cell or inhibiting a metabolic pathway, transcription, or translation such that cell proliferation slows or stops. Any agent that, when internalized inhibits or destroys cell growth, cell proliferation or other essential cell functions is suitable for use herein. Cytotoxic agents include ribosome inactivating proteins, small metabolic inhibitors, antisense nucleic acids, toxic drugs, such as anticancer agents, and small molecules, such as light activated porphyrins. Ribosome inactivating proteins, such as saporin, are the preferred cytotoxic protein agents for use herein and nucleic acids are the preferred non-peptide agents.

Such cytotoxic agents, include, but are not limited to, saporin, the ricins, abrin and other RIPs, *Pseudomonas* exotoxin, inhibitors of DNA, RNA or protein synthesis, including antisense nucleic acids and other metabolic inhibitors that are known to those of skill in this art. Saporin is preferred, but other suitable RIPs include, but are not limited to, ricin, ricin A chain, maize RIP, gelonin, diphtheria toxin and diphtheria toxin A chain (*see, e.g.*, U.S. Patent No. 4,675,382), trichosanthin, tritin, pokeweed antiviral protein (PAP), mirabilis antiviral protein (MAP), Dianthins 32 and 30, abrin, monordin, bryodin, shiga, cytotoxically active fragments of cytoxins and others known to those of skill in this art (*see, e.g.*, Barbieri et al. (1982) *Cancer Surveys 1*:489-520 and European published patent application No. 0466 222, incorporated herein by reference, which provide lists of numerous RIPs and their sources; see. also, U.S. Patent No. 5,248,608).

The selected cytotoxic agent is, if necessary, derivatized to produce a group reactive with a cysteine on the selected HBEGF. If derivatization results in a mixture of reactive species, a mono-derivatized form of the cytotoxic agent can be isolated and then conjugated to the selected HBEGF.

2. Ribosome inactivating proteins

Ribosome-inactivating-proteins (RIPs), which include ricin, abrin and saporin, are plant proteins that catalytically inactivate eukaryotic ribosomes. RIPs inactivate ribosomes by interfering with the protein elongation step of protein synthesis.

For example, the RIP saporin (hereinafter also referred to as SAP) has been shown to enzymatically inactivate 60S ribosomes by cleavage of the n-glycosidic bond of the adenine at position 4324 in the rat 28S ribosomal RNA (rRNA). Some RIPs, such as the toxins abrin and ricin, contain two constituent chains: a cell-binding chain that mediates binding to cell surface receptors and internalization of the molecule; and an enzymatically active chain responsible for protein synthesis inhibitory activity. Such RIPs are type II RIPs. Other RIPs, such as the saporins, are single chains and are designated type I RIPs. Because such RIPs lack a cell-binding chain, they far less toxic to whole cells than the RIPs that have two chains.

Several structurally related saporins have been isolated from seeds and leaves of the plant *Saponaria officinalis* (soapwort). Among these, SAP-6 is the most active and abundant, representing 7% of total seed proteins. Saporin is very stable, has a high isoelectric point, does not contain carbohydrates, and is resistant to denaturing agents, such as sodium dodecyl sulfate (SDS), and a variety of proteases. The amino acid sequences of several saporin-6 isoforms from seeds are known and there appear to be families of saporin RIPs differing in a few amino acid residues. Because saporin is a type I RIP, it does not possess a cell-binding chain. Consequently, its toxicity to whole cells is much lower than the other toxins, such as ricin and abrin. When internalized by eukaryotic cells, however, its cytotoxicity is 100- to 1000-fold more potent than ricin A chain.

Saporin is preferred herein. The saporin polypeptides include any of the isoforms of saporin that may be isolated from *Saponaria officinalis* or related species or modified form that retain cytotoxic activity. Such modified forms have amino acid substitutions, deletions, insertions or additions but still express substantial ribosome-inactivating activity. Purified preparations of saporin are frequently observed to include several molecular isoforms of the protein. It is understood that differences in amino acid sequences can occur in saporin from different species as well as between saporin molecules from individual organisms of the same species. In particular, such modified saporin may be produced by modifying the DNA encoding the protein (*see. e.g.*, published International PCT Application WO 93/25688 (Serial No. PCT/US93/05702), which is a continuation-in-part of United States Application Serial No. 07/901,718; see. also, copending U.S. Patent Application No. 07/885,242 filed May 20, 1992, and Patent No. 1231914, granted in Italy on January 15, 1992) by altering one or more amino acids or deleting or inserting one or more amino acids, such as a cysteine that may render it easier to conjugate to HBEGF or other cell surface binding protein. Any such protein, or portion thereof, that, when conjugated to HBEGF as described herein, exhibits

cytotoxicity in standard *in vitro* or *in vivo* assays within at least about an order of magnitude of the saporin conjugates described herein is contemplated for use herein.

Thus, the SAP used herein includes any protein that is isolated from natural sources or that is produced by recombinant expression (*see, e.g.*, copending published International PCT Application WO 93/25688 (Serial No. PCT/US93/05702), which is a continuation-in-part of United States Application Serial No. 07/901,718, filed June 16, 1992; see, also Example 1, below).

Some of the DNA molecules provided herein encode saporin that has substantially the same amino acid sequence and ribosome-inactivating activity as that of preferred saporin-6 (SO-6), including any of four isoforms, which have heterogeneity at amino acid positions 48 and 91 (*see, e.g.*, Maras et al., *Biochem. Internat. 21*:631-638, 1990, and Barra et al., *Biotechnol. Appl. Biochem. 13*:48-53, 1991; GB Patent 2,216,891 B and EP Patent 89306106; and SEQ ID NOS. 8-12). Other suitable saporin polypeptides include other members of the multi-gene family coding for isoforms of saporin-type ribosome-inactivating proteins including SO-1 and SO-3 (Fordham-Skelton et al., *Mol. Gen. Genet. 221*:134-138, 1990), SO-2 (*see, e.g.*, U.S. Application Serial No. 07/885,242, which corresponds to GB 2,216,891; *see, also*, Fordham-Skelton et al., *Mol. Gen. Genet. 229*:460-466, 1991), SO-4 (*see, e.g.*, GB 2,194,241 B; *see, also*, Lappi et al., *Biochem. Biophys. Res. Commun. 129*:934-942, 1985) and SO-5 (*see, e.g.*, GB 2,194,241 B; *see, also*, Montecucchi et al., *Int. J. Peptide Protein Res. 33*:263-267, 1989).

The saporin polypeptides exemplified herein include those having substantially the same amino acid sequence as those listed in SEQ ID NOS. 8-12. The isolation and expression of the DNA encoding these proteins is described in the Examples.

The saporin polypeptides include any of the isoforms of saporin that may be isolated from *Saponaria officinalis* or related species or modified forms that retain cytotoxic activity. In particular, such modified saporin may be produced by modifying the DNA encoding the protein (*see, e.g.*, International PCT Application Serial No. PCT/US93/05702, filed on June 14, 1993, and United States Application Serial No. 07/901,718; *see, also*, copending U.S. Patent Application No. 07/885,242 filed May 20, 1992, and Italian Patent No. 1,231,914) by altering one or more amino acids or deleting or inserting one or more amino acids. Any such protein, or portion thereof, that exhibits cytotoxicity in standard *in vitro* or *in vivo* assays within at least about an order of magnitude of the saporin conjugates described herein is contemplated for use herein.

b. Nucleic acids encoding other ribosome-inactivating proteins and cytocides

In addition to saporin discussed above, other cytocides that inhibit protein synthesis are useful in the present invention. The gene sequences for these cytocides may be isolated by standard methods, such as PCR, probe hybridization of genomic or cDNA libraries, antibody screenings of expression libraries, or obtain clones from commercial or other sources. The DNA sequences of many of these cytocides are well known, including ricin A chain (Genbank Accession No. X02388); maize ribosome-inactivating protein (Genbank Accession No. L26305); gelonin (Genbank Accession No. L12243; PCT Application WO 92/03155; U.S. Patent No. 5,376,546; diphtheria toxin (Genbank Accession No. K01722); trichosanthin (Genbank Accession No. M34858); tritin (Genbank Accession No. D13795); pokeweed antiviral protein (Genbank Accession No. X78628); mirabilis antiviral protein (Genbank Accession No. D90347); dianthin 30 (Genbank Accession No. X59260); abrin (Genbank Accession No. X55667); shiga (Genbank Accession No. M19437) and *Pseudomonas* exotoxin (Genbank Accession Nos. K01397, M23348).

DNA encoding SAP or any cytotoxic agent may be used in the recombinant methods provided herein. In instances in which the cytotoxic agent does not contain a cysteine residue, such as instances in which DNA encoding SAP is selected, the DNA may be modified to include a cysteine codon. The codon may be inserted into any locus that does not reduce or reduces by less than about one order of magnitude the cytotoxicity of the resulting protein. Such locus may be determined empirically by modifying the protein and testing it for cytotoxicity in an assay, such as a cell-free protein synthesis assay. The preferred loci in SAP for insertion of the cysteine residue is at or near the N-terminus (within about 20 residues, preferably 10 residues, of the N-terminus).

3. Expression of cytotoxic agents

Host organisms include those organisms in which recombinant production of heterologous proteins have been carried out and in which the cytotoxic agent, such as saporin is not toxic or of sufficiently low toxicity to permit expression before cell death. Presently preferred host organisms are strains of bacteria. Most preferred host organisms are strains of *E. coli*, particularly, BL21(DE3) cells (Novagen, Madison, WI).

The DNA encoding the cytotoxic agent, such as saporin protein, is introduced into a plasmid in operative linkage to an appropriate promoter for expression of polypeptides in a selected host organism. The presently preferred saporin proteins are saporin proteins that have been modified by addition of a Cys residue or

replacement of a non-essential residue at or near the amino- or carboxyl terminus of the saporin with Cys. Saporin, such as that of SEQ ID NO. 8 has been modified by insertion of Met-Cys residue at the N-terminus and has also been modified by replacement of the Ile or Asn residue at positions 4 and 10, respectively (see Example 4). The DNA fragment encoding the saporin may also include a protein secretion signal that functions in the selected host to direct the mature polypeptide into the periplasm or culture medium. The resulting saporin protein can be purified by methods routinely used in the art, including, methods described hereinafter in the Examples.

Methods of transforming suitable host cells, preferably bacterial cells, and more preferably *E. coli* cells, as well as methods applicable for culturing said cells containing a gene encoding a heterologous protein, are generally known in the art. See, for example, Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

The DNA construct encoding the saporin protein is introduced into the host cell by any suitable means, including, but not limited to transformation employing plasmids, bacterial phage vectors, transfection, electroporation, lipofection, and the like. The heterologous DNA can optionally include sequences, such as origins of replication that allow for the extrachromosomal maintenance of the saporin-containing plasmid, or can be designed to integrate into the genome of the host (as an alternative means to ensure stable maintenance in the host).

Positive transformants can be characterized by Southern blot analysis (Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY) for the site of DNA integration; Northern blots for inducible-promoter-responsive saporin gene expression; and product analysis for the presence of saporin-containing proteins in either the cytoplasm, periplasm, or the growth media.

Once the saporin-encoding DNA fragment has been introduced into the host cell, the desired saporin-containing protein is produced by subjecting the host cell to conditions under which the promoter is induced, whereby the operatively linked DNA is transcribed. In a preferred embodiment, such conditions are those that induce expression from the *E. coli* lac operon. The plasmid containing the DNA encoding the saporin-containing protein also includes the lac operator (O) region within the promoter and may also include the lac I gene encoding the lac repressor protein (*see, e.g.*, Muller-Hill et al. (1968) *Proc. Natl. Acad. Sci. USA 59*:1259-12649). The lac repressor represses the expression from the lac promoter until induced by the addition of IPTG in

an amount sufficient to induce transcription of the DNA encoding the saporin-containing protein.

The expression of saporin in *E. coli* is, thus accomplished in a two-stage process. In the first stage, a culture of transformed *E. coli* cells is grown under conditions in which the expression of the saporin-containing protein within the transforming plasmid, preferably encoding a saporin, such as described in Example 4, is repressed by virtue of the lac repressor. In this stage cell density increases. When an optimum density is reached, the second stage commences by addition of IPTG, which prevents binding of repressor to the operator thereby inducing the lac promoter and transcription of the saporin-encoding DNA.

In a preferred embodiment, the promoter is the T7 RNA polymerase promoter, which may be linked to the lac operator and the *E. coli* host strain includes DNA encoding T7 RNA polymerase operably linked to the lac operator and a promoter, preferably the lacUV5 promoter. The presently preferred plasmid is pET 11a (Novagen, Madison, WI), which contains the T7lac promoter, T7 terminator, the inducible *E. coli* lac operator, and the lac repressor gene. The plasmid pET 15b (Novagen, Madison, WI), which contains a His-Tag™ leader sequence (Seq. ID NO. 23) for use in purification with a His column and a thrombin cleavage site that permits cleavage following purification over the column, the T7-lac promoter region and the T7 terminator, has been used herein for expression of saporin. Addition of IPTG induces expression of the T7 RNA polymerase and the T7 promoter, which is recognized by the T7 RNA polymerase.

Transformed strains, which are of the desired phenotype and genotype, are grown in fermentors by suitable methods well known in the art. In the first, or growth stage, expression hosts are cultured in defined minimal medium lacking the inducing condition, preferably IPTG. When grown in such conditions, heterologous gene expression is completely repressed, which allows the generation of cell mass in the absence of heterologous protein expression. Subsequent to the period of growth under repression of heterologous gene expression, the inducer, preferably IPTG, is added to the fermentation broth, thereby inducing expression of any DNA operatively linked to an IPTG-responsive promoter (a promoter region that contains lac operator). This last stage is the induction stage.

The resulting saporin-containing protein can be suitably isolated from the other fermentation products by methods routinely used in the art, *e.g.*, using a suitable affinity column as described in the Examples; precipitation with ammonium sulfate; gel filtration; chromatography, preparative flat-bed iso-electric focusing; gel

electrophoresis, high performance liquid chromatography (HPLC); and the like. A method for isolating saporin is provided in Example 1 (see, also Lappi et al. ((1985) *Biochem. Biophys. Res. Commun., 129*:934-942). The expressed saporin protein is isolated from either the cytoplasm, periplasm, or the cell culture medium (see, discussion below and *see, e.g.*, Example 3 for preferred methods and saporin proteins).

4. Porphyrins

Porphyrins are well known light activatable toxins that can be readily cross-linked to proteins (*see, e.g.*, U.S. Patent No. 5,257,970; U.S. Patent No. 5,252,720; U.S. Patent No. 5,238,940; U.S. Patent No. 5,192,788; U.S. Patent No. 5,171,749; U.S. Patent No. 5,149,708; U.S. Patent No. 5,202,317; U.S. Patent No. 5,217,966; U.S. Patent No. 5,053,423; U.S. Patent No. 5,109,016; U.S. Patent No. 5,087,636; U.S. Patent No. 5,028,594; U.S. Patent No. 5,093,349; U.S. Patent No. 4,968,715; U.S. Patent No. 4,920,143 and International Application WO 93/02192).

Porphyrins are conjugated to proteins by direct, covalent bonds using, for example, a carbodiimide. Linkage may be effected by treatment of HBEGF by 1-ethyl-3-3-dimethylamino propyl) carbo diimide in the presence of a reaction medium such as DMSO. For other methods see U.S. Patent No. 4,968,715. The porphyrin HBEGF conjugates may be administered topically or systemically. Actuation of the porphyrin is by irradiating light chosen to match the maximum absorbance of the porphyrin-type photosensitizer.

5. Nucleic acids for targeted delivery

The conjugates provided herein are also designed to deliver nucleic acids to targeted cells. The nucleic acids include those intended to deliver a cytotoxic signal to a cell or to modify expression of genes and thereby effect genetic therapy. Examples of nucleic acids include antisense RNA, DNA, ribozymes and oligonucleotides that bind proteins. The nucleic acids can also include RNA trafficking signals, such as viral packaging sequences (*see, e.g.*, Sullenger et al. (1994) *Science 262*:1566-1569). The nucleic acids also include DNA molecules that encode intact genes or that encode proteins useful for gene therapy or for effecting cell cytotoxicity. Especially of interest are DNA molecules that encode an enzyme that results in cell death or renders a cell susceptible to cell death upon the addition of another product. For example, saporin is an enzyme that cleaves rRNA and inhibits protein synthesis. Other enzymes that inhibit protein synthesis are especially well suited for the present invention. Other enzymes may be used where the enzyme activates a compound with little or no cytotoxicity into a toxic product.

DNA (or RNA) that may be delivered to a cell to effect genetic therapy includes DNA that encodes tumor-specific cytotoxic molecules, such as tumor necrosis factor, viral antigens and other proteins to render a cell susceptible to anti-cancer agents, and DNA encoding genes, such as the such as the defective gene (CFTR) associated with cystic fibrosis (*see, e.g.*, International Application WO 93/03709, which is based on U.S. Application Serial No. 07/745,900; and Riordan et al. (1989) *Science 245*:1066-1073), to replace defective genes.

Nucleic acids and oligonucleotides for use as described herein can be synthesized by any method known to those of skill in this art (*see, e.g.*, Wo 93/01286, which is based on U.S. Application Serial No. 07/723,454; U.S.. Patent No. 5,218,088; U.S. Patent No. 5,175,269; U.S. Patent No. 5,109,124). Identification of oligonucleotides and ribozymes for use as antisense agents as well as selection of DNA encoding genes for targeted delivery for genetic therapy, is well within the skill in this art. For example, the desirable properties, lengths and other characteristics of such oligonucleotides are well known. Antisense oligonucleotides are designed to resist degradation by endogenous nucleolytic enzymes and include, but are not limited to: phosphorothioate, methylphosphonate, sulfone, sulfate, ketyl, phosphorodithioate, phosphoramidate, phosphate esters, and other such linkages (*see, e.g.*, Agrwal et al. (1987) *Tetrehedron Lett. 28*:3539-3542; Miller et al. (1971) *J. Am. Chem. Soc. 93*:6657-6665; Stec et al. (1985) *Tetrehedron Lett. 26*:2191-2194; Moody et al. (1989) *Nucl. Acids Res. 12*:4769-4782; Uznanski et al. (1989); *Nucl. Acids Res.* Letsinger et al. (1984) *Tetrahedron 40*:137-143; Eckstein (1985) *Annu. Rev. Biochem. 54*:367-402; Eckstein (1989) *Trends Biol. Sci. 14*:97-100; Stein (1989) In: *Oligodeoxynucleotides. Antisense Inhibitors of Gene Expression*, Cohen, Ed, Macmillan Press, London, pp. 97-117; Jager et al. (1988) *Biochemistry 27*:7237-7246).

a. Antisense nucleotides

Antisense nucleotides are oligonucleotides that bind in a sequence-specific manner to nucleic acids, such as mRNA or DNA. When bound to mRNA that has complementary sequences, antisense prevents translation of the mRNA (*see, e.g.*, U.S. Patent No. 5,168,053 to Altman et al., U.S. Patent No. 5,190,931 to Inouye, U.S. Patent No. 5,135,917 to Burch; U.S. Patent No. 5,087,617 to Smith and Clusel et al. (1993) *Nucl. Acids Res. 21*:3405-3411, which describes dumbbell antisense oligonucleotides). Triplex molecules refer to single DNA strands that bind duplex DNA forming a colinear triplex molecule and thereby prevent transcription (*see, e.g.*, U.S. Patent No. 5,176,996 to Hogan et al., which describes methods for making synthetic oligonucleotides that bind to target sites on duplex DNA).

Particularly useful antisense nucleotides and triplex molecules are molecules that are complementary or bind to the sense strand of DNA or mRNA that encodes an oncogene, such as bFGF, int-2, hst-1/K-FGF, FGF-5, hst-2/FGF-6, FGF-8. Other useful antisense oligonucleotides include those that are specific for IL-8 (*see, e.g.*, U.S. Patent No. 5,241,049; and International applications WO 89/004836; WO 90/06321; WO 89/10962; WO 90/00563; and WO 91/08483, and the corresponding U.S. applications for descriptions of DNA encoding IL-8 and amino acid sequences of IL-8), which can be linked to bFGF for the treatment of psoriasis, anti-sense oligonucleotides that are specific for nonmuscle myosin heavy chain and/or c-myb (*see, e.g.*, Simons et al. (1992) *Circ. Res. 70*:835-843; WO 93/01286, which is based on U.S. application Serial No. 07/723,454: LeClerc et al. (1991) *J. Am. Coll. Cardiol. 17 (2 Suppl. A)*:105A; Ebbecke et al. (1992) *Basic Res. Cardiol. 87*:585-591), which can be targeted by an FGF to inhibit smooth muscle cell proliferation, such as that following angioplasty and thereby prevent restenosis or inhibit viral gene expression in transformed or infected cells.

b. Ribozymes

A ribozyme is an RNA molecule that specifically cleaves RNA substrates, such as messenger RNA, and thus inhibits or interferes with cell growth or expression. There are at least five classes of ribozymes that are known that are involved in the cleavage and/or ligation of RNA chains. Ribozymes can be targeted to any RNA transcript and can catalytically cleave such transcript (*see, e.g.*, U.S. Patent No. 5,272,262; U.S. Patent No. 5,144,019; and U.S. Patent Nos. 5,168,053, 5,180,818, 5,116,742 and 5,093,246 to Cech et al., which described ribozymes and methods for production thereof). Any such ribozyme may be linked to the growth factor for delivery to HBEGF-receptor bearing cells.

The ribozymes may be delivered to the targeted cells, such as tumor cells that express a receptor to which HBEGF binds and upon binding is internalized, as DNA encoding the ribozyme linked to a eukaryotic promoter, such as a eukaryotic viral promoter, generally a later promoter, such that upon introduction into the nucleus, the ribozyme will be directly transcribed. In such instances, the construct will also include a nuclear translocation sequence (NTS; see Table 2, below), generally as part of the growth factor or as part of a linker between the growth factor and linked DNA.

c. Nucleic acids encoding therapeutic products

Among the DNA that encodes therapeutic products contemplated for use is DNA encoding correct copies of defective genes, such as the defective gene (CFTR) associated with cystic fibrosis (*see, e.g.*, International Application WO 93/03709, which

is based on U.S. Application Serial No. 07/745,900; and Riordan et al. (1989) *Science 245*:1066-1073), and anticancer agents, such as tumor necrosis factors, and cytotoxic agents, such as saporin. The conjugate should include an NTS. If the conjugate is designed such that the HBEGF and linked DNA is cleaved in the cytoplasm, then the NTS should be included in a portion of the linker that remains bound to the DNA, so that, upon internalization, the conjugate will be trafficked to the nucleus. The nuclear translocation sequence (NTS) may be a heterologous sequence or a may be derived from the selected growth factor.

d. Other nucleic acids

Extracellular protein binding oligonucleotides refer to oligonucleotides that specifically bind to proteins. Small nucleotide molecules refer to nucleic acids that target a receptor site.

e. Coupling of nucleic acids to proteins

To effect chemical conjugation herein, the HBEGF protein is linked to the nucleic acid either directly or via one or more linkers. Methods for conjugating nucleic acids, at the 5' ends, 3' ends and elsewhere, to the amino and carboxyl termini and other sites in proteins are known to those of skill in the art (for a review see *e.g.*, Goodchild, (1993) In: *Perspectives in Bioconjugate Chemistry*, Mears, Ed., American Chemical Society, Washington, D.C. pp. 77-99). For example, proteins have been linked to nucleic acids using ultraviolet irradiation (Sperling et al. (1978) *Nucleic Acids Res. 5*:2755-2773; Fiser et al. (1975) *FEBS Lett. 52*:281-283), bifunctional chemicals (Bäumert et al. (1978) *Eur. J. Biochem. 89*:353-359; and Oste et al. (1979) *Mol. Gen. Genet. 168*:81-86) photochemical cross-linking (Vanin et al. (1981) *FEBS Lett. 124*:89-92; Rinke et al. (1980) *J.Mol.Biol. 137*:301-314; Millon et al. (1980); *Eur. J. Biochem. 110*:485-454).

In particular, the reagents (N-acetyl-N'-(p-glyoxylylbenzolyl) cystamine and 2-iminothiolane have been used to couple DNA to proteins, such as α 2macroglobulin (α_2M) via mixed disulfide formation (see, Cheng et al. (1983) *Nucleic Acids Res. 11*:659-669). N-acetyl-N'-(p-glyoxylylbenzolyl)cystamine reacts specifically with nonpaired guanine residues and, upon reduction, generates a free sulfhydryl group. 2-Iminothiolane reacts with proteins to generate sulfhydryl groups that are then conjugated to the derivatized DNA by an intermolecular disulfide interchange reaction. Any linkage may be used provided that, upon internalization of the conjugate the targeted nucleic acid is active. Thus, it is expected that cleavage of the linkage may be necessary, although it is contemplated that for some reagents, such as DNA encoding

ribozymes linked to promoters or DNA encoding therapeutic agents for delivery to the nucleus, such cleavage may not be necessary.

Thiol linkages can be readily formed using heterobifunctional reagents. Amines have also been attached to the terminal 5' phosphate of unprotected oligonucleotides or nucleic acids in aqueous solutions by reacting the nucleic acid with a water-soluble carbodiimide, such as 1-ethyl-3'[3-dimethylaminopropyl]carbodiimide (EDC) or N-ethyl-N'(3-dimethylaminopropylcarbodiimidehydrochloride (EDCI), in imidazole buffer at pH 6 to produce the 5'phosphorimidazolide. Contacting the 5'phosphorimidazolide with amine-containing molecules, such as HBEGF, and ethylenediamine, results in stable phosphoramidates (*see, e.g.*, Chu et al. (1983) *Nucleic Acids Res. 11*:6513-6529; and WO 88/05077 in which the U.S. is designated). In particular, a solution of DNA is saturated with EDC, at pH 6 and incubated with agitation at 4^{o} C overnight. The resulting solution is then buffered to pH 8.5 by adding, for example about 3 volumes of 100 mM citrate buffer, and adding about 5 μg - 20 μg of an HBEGF, and agitating the resulting mixture at 4^{o} C for about 48 hours. The unreacted protein may be removed from the mixture by column chromatography using, for example, Sephadex G75 (Pharmacia) using 0.1 M ammonium carbonate solution, pH 7.0 as an eluting buffer. The isolated conjugate may be lyophilized and stored until used.

U.S. Patent No. 5,237,016 provides methods for preparing nucleotides that are bromacetylated at their 5' termini and reacting the resulting oligonucleotides with thiol groups. Oligonucleotides derivatized at their 5'-termini bromoacetyl groups can be prepared by reacting 5'-aminohexyl-phosphoramidate oligonucleotides with bromoacetic acid-N-hydroxysuccinimide ester as described in U.S. Patent No. 5,237,016. U.S. Patent No. 5,237,016 also describes methods for preparing thiol-derivatized nucleotides, which can then be reacted with thiol groups on the selected growth factor. Briefly, thiol-derivatized nucleotides are prepared using a 5'-phosphorylated nucleotide in two steps: (1) reaction of the phosphate group with imidazole in the presence of a diimide and displacement of the imidazole leaving group with cystamine in one reaction step; and reduction of the disulfide bond of the cystamine linker with dithiothreitol (see, also, Orgel et al. (1986) *Nucl. Acids Res. 14*:651, which describes a similar procedure). The 5'-phosphorylated starting oligonucleotides can be prepared by methods known to those of skill in the art (*see, e.g.*, Maniatis et al. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York, p. 122).

The antisense oligomer or nucleic acid, such as a methylphosphonate oligonucleotide (MP-oligomer), may be derivatized by reaction with SPDP or SMPB.

The resulting MP-oligomer may be purified by HPLC and then coupled to HBEGF, which may be modified by replacement of one or more non-essential cysteine residues, as described above. The MP-oligomer (about 0.1 μM) is dissolved in about 40-50 μl of 1:1 acetonitrile/water to which phosphate buffer (pH 7.5, final concentration 0.1 M) and a 1 mg MP-oligomer in about 1 ml phosphate buffered saline is added. The reaction is allowed to proceed for about 5-10 hours at room temperature and is then quenched with about 15 μL 0.1 iodoacetamide. The HBEGF-oligonucleotide conjugates can be purified on heparin sepharose Hi Trap columns (1 ml, Pharmacia) and eluted with a linear or step gradient. The conjugate should elute in 0.6 M NaCl.

f. Nucleic acids encoding cytocides

A cytocide-encoding agent is a nucleic acid molecule (DNA or RNA) that, upon internalization by a cell, and subsequent transcription and/or translation into a cytocidal agent, is cytotoxic to a cell or inhibits cell growth by inhibiting protein synthesis.

Cytocides include saporin, the ricins, abrin and other ribosome-inactivating proteins, *Pseudomonas* exotoxin, diptheria toxin, angiogenin, tritin, dianthins 32 and 30, momordin, pokeweed antiviral protein, mirabilis antiviral protein, bryodin, angiogenin, and shiga exotoxin, as well as other cytocides that are known to those of skill in the art.

Especially of interest are DNA molecules that encode an enzyme that results in cell death or renders a cell susceptible to cell death upon the addition of another product. For example, saporin, a preferred cytocide, is an enzyme that cleaves rRNA and inhibits protein synthesis. Other enzymes that inhibit protein synthesis are especially well suited for use in the present invention. In addition, enzymes may be used where the enzyme activates a compound with little or no cytotoxicity into a toxic product that inhibits protein synthesis.

In addition to saporin discussed above, other cytocides that inhibit protein synthesis are useful in the present invention. The gene sequences for these cytocides may be isolated by standard methods, such as PCR, probe hybridization of genomic or cDNA libraries, antibody screenings of expression libraries, or obtain clones from commercial or other sources. The DNA sequences of many of these cytocides are well known, including ricin A chain (Genbank Accession No. X02388); maize ribosome-inactivating protein (Genbank Accession No. L26305); gelonin (Genbank Accession No. L12243; PCT Application WO 92/03155; U.S. Patent No. 5,376,546; diphtheria toxin (Genbank Accession No. K01722); trichosanthin (Genbank Accession No. M34858); tritin (Genbank Accession No. D13795); pokeweed antiviral

protein (Genbank Accession No. X78628); mirabilis antiviral protein (Genbank Accession No. D90347); dianthin 30 (Genbank Accession No. X59260); abrin (Genbank Accession No. X55667); shiga (Genbank Accession No. M19437) and *Pseudomonas* exotoxin (Genbank Accession Nos. K01397, M23348).

In the case of cytotocidal molecules such as the ribosome-inactivating proteins, very few molecules may need be present for cell killing. Indeed, only a single molecule of diphtheria toxoid introduced into a cell was sufficient to kill the cell. In other cases, it may be that propagation or stable maintenance of the construct is necessary to attain sufficient numbers or concentrations of the gene product for effective gene therapy. Examples of replicating and stable eukaryotic plasmids are found in the scientific literature.

In general, constructs will also contain elements necessary for transcription and translation. If the cytocide-encoding agent is DNA, then it must contain a promoter. The choice of the promoter will depend upon the cell type to be transformed and the degree or type of control desired. Promoters can be constitutive or active in any cell type, tissue specific, cell specific, event specific or inducible. Cell-type specific promoters and event type specific promoters are preferred. Examples of constitutive or nonspecific promoters include the SV40 early promoter (U.S. Patent No. 5,118,627), the SV40 late promoter (U.S. Patent No. 5,118,627), CMV early gene promoter (U.S. Patent No. 5,168,062), and adenovirus promoter. In addition to viral promoters, cellular promoters are also amenable within the context of this invention. In particular, cellular promoters for the so-called housekeeping genes are useful.

Tissue specific promoters are particularly useful when a particular tissue type is to be targeted for transformation. By using one of this class of promoters, an extra margin of specificity can be attained. For example, when the indication to be treated is ophthalmological, either the alpha-crystalline promoter or gamma-crystalline promoter is preferred. When a tumor is the target of gene delivery, cellular promoters for specific tumor markers or promoters more active in tumor cells should be chosen. Thus, to transform prostate tumor cells the prostate-specific antigen promoter is especially useful. Similarly, the tyrosinase promoter or tyrosinase-related protein promoter is a preferred promoter for melanoma treatment. For B lymphocytes, the immunoglobulin variable region gene promoter, for T lymphocytes, the TCR receptor variable region promoter, for helper T lymphocytes, the CD4 promoter, for liver, the albumin promoter, are but a few examples of tissue specific promoters. In certain applications, such as treatment of restenosis, a promoter for myosin light chain specific

for smooth muscle cells is preferred. Many other examples of tissue specific promoters are readily available to one skilled in the art.

Inducible promoters may also be used. These promoters include the MMTV LTR (PCT WO 91/13160), which is inducible by dexamethasone, metallothionein, which is inducible by heavy metals, and promoters with cAMP response elements, which are inducible by cAMP. By using an inducible promoter, the nucleic acid may be delivered to a cell and will remain quiescent until the addition of the inducer. This allows further control on the timing of production of the therapeutic gene.

Event-type specific promoters are active only upon the occurrence of an event, such as tumorigenecity or viral infection. The HIV LTR is a well known example of an event-specific promoter. The promoter is inactive unless the *tat* gene product is present, which occurs upon viral infection.

Additionally, promoters that are coordinately regulated with a particular cellular gene may be used. For example, promoters of genes that are coordinately expressed when a particular HBEGF receptor gene is expressed may be used. Then, the nucleic acid will be transcribed when the HBEGF receptor is expressed. This type of promoter is especially useful when one knows the pattern of HBEGF receptor expression in a particular tissue, so that specific cells within that tissue may be killed upon transcription of a cytotoxic agent gene without affecting the surrounding tissues.

Alternatively, cytocide gene products may be noncytotoxic but activate a compound, which is endogenously produced or exogenously applied, from a nontoxic form to a toxic product that inhibits protein synthesis.

The construct must contain the sequence that binds to the nucleic acid binding domain, if the domain binds in a sequence specific manner. As described below, the target nucleotide sequence may be contained within the coding region of the cytocide, in which case, no additional sequence need be incorporated. It may be desirable to have multiple copies of target sequence. If the target sequence is coding sequence, the additional copies must be located in non-coding regions of the cytocide-encoding agent. The target sequences of the nucleic acid binding domains are typically generally known. The target sequence may be readily determined, in any case. Techniques are generally available for establishing the target sequence (*e.g.*, *see* PCT Application WO 92/05285 and U.S. Serial No. 586,769).

Specificity of delivery is achieved by coupling a nucleic acid binding domain to a receptor-binding internalized ligand, either by chemical conjugation or by constructing a fusion protein. Linkers as described above may be used. The receptor-

binding internalized ligand part confers specificity of delivery in a cell-specific manner. The choice of the receptor-binding internalized ligand to use will depend upon the receptor expressed by the target cells. The receptor type of the target cell population may be determined by conventional techniques such as antibody staining, PCR of cDNA using receptor-specific primers, and biochemical or functional receptor binding assays. It is preferable that the receptor be cell type specific or have increased expression or activity (*i.e.*, higher rate of internalization) within the target cell population.

The nucleic acid binding domain can be of two types, non-specific in its ability to bind nucleic acid, or highly specific so that the amino acid residues bind only the desired nucleic acid sequence. Nonspecific binding proteins, polypeptides, or compounds are generally polycations or highly basic. Lys and Arg are the most basic of the 20 common amino acids; proteins enriched for these residues are candidates for nucleic acid binding domains. Examples of basic proteins include histones, protamines, and repeating units of lysine and arginine. Poly-L-lysine is a well-used nucleic acid binding domain (*see* U.S. Patent Nos. 5,166,320 and 5,354,844). Other polycations, such as spermine and spermidine, may also be used to bind nucleic acids. By way of example, the sequence-specific proteins including Sp-1, AP-1, myoD and the *rev* gene product from HIV may be used. Specific nucleic acid binding domains can be cloned in tandem, individually, or multiply to a desired region of the receptor-binding internalized ligand of interest. Alternatively, the domains can be chemically conjugated to each other.

The corresponding response elements that bind sequence-specific domains are incorporated into the construct to be delivered. Complexing the cytocidal-encoding agent to the receptor-binding internalized ligand/nucleic acid binding domain allows specific binding of response element to the nucleic acid binding domain. Even greater specificity of binding may be achieved by identifying and using the minimal amino acid sequence that binds to the cytocidal-encoding agent of interest. For example, phage display methods can be used to identify amino acids residues of varying length that will bind to specific nucleic acid sequences with high affinity. (*See* U.S. Patent No. 5,223,409.) The peptide sequence can then be cloned into the receptor-binding internalized ligand as a single copy or multiple copies. Alternatively, the peptide may be chemically conjugated to the receptor-binding internalized ligand. Incubation of the cytocide-encoding agent with the conjugated proteins will result in a specific binding between the two.

These complexes may be used to deliver nucleic acids that encode saporin or other cytocidal proteins into cells that have appropriate receptors that are expressed, over-expressed or more active in internalization upon binding. The cytocide gene is cloned downstream of a mammalian promoter such as SV40, CMV, TK or Adenovirus promoter. As described above, promoters of interest may be active in any cell type, active only in a tissue-specific manner, such as α-crystalline or tyrosinase, event specific or inducible, such as the MMTV LTR.

Receptor-binding internalized ligands are prepared as discussed by any suitable method, including recombinant DNA technology, isolation from a suitable source, purchase from a commercial source, or chemical synthesis. The selected linker or linkers is (are) linked to the receptor-binding internalized ligands by chemical reaction, generally relying on an available thiol or amine group on the receptor-binding internalized ligands. Heterobifunctional linkers are particularly suited for chemical conjugation. Alternatively, if the linker is a peptide linker, then the receptor-binding internalized ligands, linker and nucleic acid binding domain can be expressed recombinantly as a fusion protein.

HBEGF may be isolated from a suitable source or may be produced using recombinant DNA methodology, discussed below. To effect chemical conjugation herein, the growth factor protein is conjugated generally via a reactive amine group or thiol group to the nucleic acid binding domain directly or through a linker to the nucleic acid binding domain. The growth factor protein is conjugated either via its N-terminus, C-terminus, or elsewhere in the polypeptide. In preferred embodiments, the growth factor protein is conjugated via a reactive cysteine residue to the linker or to the nucleic acid binding domain. The growth factor can also be modified by addition of a cysteine residue, either by replacing a residue or by inserting the cysteine, at or near the amino or carboxyl terminus, within about 20, preferably 10 residues from either end, and preferably at or near the amino terminus.

In certain embodiments, the heterogeneity of preparations may be reduced by mutagenizing the growth factor protein to replace reactive cysteines, leaving, preferably, only one available cysteine for reaction. The growth factor protein is modified by deleting or replacing a site(s) on the growth factor that causes the heterogeneity. Such sites are typically cysteine residues that, upon folding of the protein, remain available for interaction with other cysteines or for interaction with more than one cytotoxic molecule per molecule of heparin-binding growth factor peptide. Thus, such cysteine residues do not include any cysteine residue that are required for proper folding of the growth factor or for retention of the ability to bind to

a growth factor receptor and internalize. For chemical conjugation, one cysteine residue that, in physiological conditions, is available for interaction, is not replaced because it is used as the site for linking the cytotoxic moiety. The resulting modified heparin-binding growth factor is conjugated with a single species of cytotoxic conjugate.

Alternatively, the contribution of each cysteine to the ability to bind to HBEGF receptors may be determined empirically. Each cysteine residue may be systematically replaced with a conservative amino acid change (see Table 1, above) or deleted. The resulting mutein is tested for the requisite biological activity: the ability to bind to growth factor receptors and internalize linked nucleic acid binding domain and agents. If the mutein retains this activity, then the cysteine residue is not required. Additional cysteines are systematically deleted and replaced and the resulting muteins are tested for activity. Each of the remaining cysteine residues may be systematically deleted and/or replaced by a serine residue or other residue that would not be expected to alter the structure of the protein. The resulting peptide is tested for biological activity. If the cysteine residue is necessary for retention of biological activity it is not deleted; if it not necessary, then it is preferably replaced with a serine or other residue that should not alter the secondary structure of the resulting protein. In this manner the minimum number and identity of the cysteines needed to retain the ability to bind to a heparin-binding growth factor receptor and internalize may be determined. It is noted, however, that modified or mutant heparin-binding growth factors may exhibit reduced or no proliferative activity, but may be suitable for use herein, if they retain the ability to target a linked cytotoxic agent to cells bearing receptors to which the unmodified heparin-binding growth factor binds and result in internalization of the cytotoxic moiety.

For recombinant expression using the methods described herein, up to all cysteines in the HBEGF polypeptide that are not required for biological activity can be deleted or replaced. Alternatively, for use in the chemical conjugation methods herein, all except one of these cysteines, which will be used for chemical conjugation to the cytotoxic agent, can be deleted or replaced. Each of the HBEGF polypeptides described herein have six cysteine residues. Each of the six cysteines may independently be replaced and the resulting mutein tested for the ability to bind to HBEGF receptors and to be internalized. Alternatively, the resulting mutein-encoding DNA is used as part of a construct containing DNA encoding the nucleic acid binding domain linked to the HBEGF-encoding DNA. The construct is expressed in a suitable host cell and the resulting protein tested for the ability to bind to HBEGF receptors and internalize. As long as this ability is retained the mutein is suitable for use herein.

The HBEGF monomers are preferably linked via non-essential cysteine residues to the linkers or to the targeted agent. HBEGF that has been modified by introduction of a Cys residue at or near one terminus, preferably the N-terminus is preferred for use in chemical conjugation. Methods for coupling proteins to the linkers, such as the heterobifunctional agents, or to nucleic acids, or to proteins are known to those of skill in the art and are also described herein.

Methods for chemical conjugation of proteins are known to those of skill in the art. The preferred methods for chemical conjugation depend on the selected components, but preferably rely on disulfide bond formation. To effect chemical conjugation herein, the HBEGF polypeptide is linked via one or more selected linkers or directly to the nucleic acid binding domain.

A nucleic acid binding domain is prepared for chemical conjugation. For chemical conjugation, a nucleic acid binding domain may be derivatized with SPDP or other suitable chemicals. If the binding domain does not have a Cys residue available for reaction, one can be either inserted or substituted for another amino acid. If desired, mono-derivatized species may be isolated, essentially as described.

For chemical conjugation, the nucleic acid binding domain may be derivatized or modified such that it includes a cysteine residue for conjugation to the receptor-binding internalized ligand. Typically, derivatization proceeds by reaction with SPDP. This results in a heterogeneous population. For example, nucleic acid binding domain that is derivatized by SPDP to a level of 0.9 moles pyridine-disulfide per mole of nucleic acid binding domain includes a population of non-derivatized, mono-derivatized and di-derivatized SAP. Nucleic acid binding domain proteins, which are overly derivatized with SPDP, may lose ability to bind nucleic acid because of reaction with sensitive lysines (Lambert et al., *Cancer Treat. Res. 37*:175-209, 1988). The quantity of non-derivatized nucleic acid binding domain in the preparation of the non-purified material can be difficult to judge and this may lead to errors in being able to estimate the correct proportion of derivatized nucleic acid binding domain to add to the reaction mixture.

Because of the removal of a negative charge by the reaction of SPDP with lysine, the three species, however, have a charge difference. The methods herein rely on this charge difference for purification of mono-derivatized nucleic acid binding domain by Mono-S cation exchange chromatography. The use of purified mono-derivatized nucleic acid binding domain has distinct advantages over the non-purified material. The amount of receptor-binding internalized ligand that can react with nucleic acid binding domain is limited to one molecule with the mono-derivatized material, and

it is seen in the results presented herein that a more homogeneous conjugate is produced. There may still be sources of heterogeneity with the mono-derivatized nucleic acid binding domain used here but is acceptable as long as binding to the cytocide-encoding agent is not impacted.

Because more than one amino group on the nucleic acid binding domain may react with the succinimidyl moiety, it is possible that more than one amino group on the surface of the protein is reactive. This creates potential for heterogeneity in the mono-derivatized nucleic acid binding domain. As an alternative to derivatizing to introduce a sulfhydryl, the nucleic acid binding domain can be modified by the introduction of a cysteine residue. Preferred loci for introduction of a cysteine residue include the N-terminus region, preferably within about one to twenty residues from the N-terminus of the nucleic acid binding domain. Using either methodology (reacting mono-derivatized nucleic acid binding domain or introducing a Cys residue into nucleic acid binding domain), the resulting preparations of chemical conjugates are monogenous; compositions containing the conjugates also appear to be free of aggregates. As a preferred alternative, heterogeneity can be avoided by producing a fusion protein of receptor-binding internalized ligand and nucleic acid binding domain, as described below.

Expression of DNA encoding a fusion of a receptor-binding internalized ligand polypeptide linked to the nucleic acid binding domain results in a more homogeneous preparation of cytotoxic conjugates. Aggregate formation can be reduced in preparations containing the fusion proteins by modifying the receptor-binding internalized ligand, such as by removal of nonessential cysteines, and/or the nucleic acid binding domain to prevent interactions between conjugates via free cysteines.

DNA encoding the polypeptides may be isolated, synthesized or obtained from commercial sources or prepared as described herein. Expression of recombinant polypeptides may be performed as described herein; and DNA encoding these polypeptides may be used as the starting materials for the methods herein.

As described above, DNA encoding HBEGF are described above. DNA may be prepared synthetically based on the amino acid or DNA sequence or may be isolated using methods known to those of skill in the art, such as PCR, probe hybridization of libraries, and the like or obtained from commercial or other sources.

As described herein, such DNA may then be mutagenized using standard methodologies to delete or replace any cysteine residues that are responsible for aggregate formation. If necessary, the identity of cysteine residues that contribute to aggregate formation may be determined empirically, by deleting and/or replacing a

cysteine residue and ascertaining whether the resulting growth factor with the deleted cysteine forms aggregates in solutions containing physiologically acceptable buffers and salts. Loci for insertion of cysteine residues may also be determined empirically. Generally, regions at or near (within 20, preferably 10 amino acids) the C- or, preferably, the N-terminus are preferred.

The DNA construct encoding the fusion protein can be inserted into a plasmid and expressed in a selected host, as described above, to produce a recombinant receptor-binding internalized ligand—nucleic acid binding domain conjugate. Multiple copies of the chimera can be inserted into a single plasmid in operative linkage with one promoter. When expressed, the resulting protein will then be a multimer. Typically, two to six copies of the chimera are inserted, preferably in a head to tail fashion, into one plasmid.

To produce monogenous preparations of fusion protein, HBEGF DNA is modified so that, upon expression, the resulting HBEGF portion of the fusion protein does not include any cysteines available for reaction. In preferred embodiments, DNA encoding an HBEGF polypeptide is linked to DNA encoding a nucleic acid binding domain. The DNA encoding the HBEGF polypeptide or other receptor-binding internalized ligand is modified in order to remove the translation stop codon and other transcriptional or translational stop signals that may be present and to remove or replace DNA encoding the available cysteines. The DNA is then ligated to the DNA encoding the nucleic acid binding domain polypeptide directly or via a linker region of one or more codons between the first codon of the nucleic acid binding domain and the last codon of the HBEGF. The size of the linker region may be any length as long as the resulting conjugate binds and is internalized by a target cell. Presently, spacer regions of from about one to about seventy-five to ninety codons are preferred. The order of the receptor-binding internalized ligand and nucleic acid binding domain in the fusion protein may be reversed. If the nucleic acid binding domain is N-terminal, then it is modified to remove the stop codon and any stop signals.

If the HBEGF or other ligand has been modified so as to lack mitogenic activity or other biological activities, binding and internalization may still be readily assayed by any one of the following tests or other equivalent tests. Generally, these tests involve labeling the ligand, incubating it with target cells, and visualizing or measuring intracellular label. For example, briefly, HBEGF may be fluorescently labeled with FITC or radiolabeled with ^{125}I. Fluorescein-conjugated HBEGF is incubated with cells and examined microscopically by fluorescence microscopy or confocal microscopy for internalization. When HBEGF is labeled with ^{125}I, the

labeled HBEGF is incubated with cells at 4°C. Cells are temperature shifted to 37°C and washed with 2 M NaCl at low pH to remove any cell-bound HBEGF. Label is then counted and thereby measuring internalization of HBEGF. Alternatively, the ligand can be conjugated with an nucleic acid binding domain by any of the methods described herein and complexed with a plasmid encoding saporin. As discussed below, the complex may be used to transfect cells and cytoxicity measured.

The DNA encoding the resulting receptor-binding internalized ligand—nucleic acid binding domain can be inserted into a plasmid and expressed in a selected host, as described above, to produce a monogenous preparation.

Multiple copies of the modified receptor-binding internalized ligand/nucleic acid binding domain chimera can be inserted into a single plasmid in operative linkage with one promoter. When expressed, the resulting protein will be a multimer. Typically two to six copies of the chimera are inserted, preferably in a head to tail fashion, into one plasmid. Merely by way of example, DNA encoding human bFGF- has been mutagenized using splicing by overlap extension (SOE). Each application of the SOE method uses two amplified oligonucleotide products, which have complementary ends as primers and which include an altered codon at the locus at which the mutation is desired, to produce a hybrid product. A second amplification reaction that uses two primers that anneal at the non-overlapping ends amplify the hybrid to produce DNA that has the desired alteration.

The receptor-binding internalized ligand/nucleic acid binding domain is incubated with the cytocide-encoding agent, typically a DNA molecule, to be delivered under conditions that allow binding of the nucleic acid binding domain to the agent. Conditions will vary somewhat depending on the nature of the nucleic acid binding domain, but will typically occur in 0.1M NaCl and 20 mM HEPES or other similar buffer.

The desired application is the delivery of cytotocidal agents, such as saporin, in a non-toxic form. By delivering a nucleic acid molecule capable of expressing saporin, the timing of cytotoxicity may be exquisitely controlled. For example, if saporin is expressed under the control of a tissue-specific promoter, then uptake of the complex by cells having the tissue-specific factors necessary for promoter activation will result in the killing of those cells. On the other hand, if cells taking up the complex do not have those tissue-specific factors, the cells will be spared.

Merely by way of example, test constructs have been made and tested. One construct is a chemical conjugate of bFGF and poly-L-lysine. The bFGF molecule is a variant in which the Cys residue at position 96 has been changed to a serine; thus,

only the Cys at position 78 is available for conjugation. This bFGF is called FGF2-3. The poly-L-lysine was derivatized with SPDP and coupled to FGF2-3. This FGF2-3/poly-L-lysine conjugate was used to deliver a plasmid able to express the β-galactosidase gene.

The ability of a construct to bind nucleic acid molecules may be conveniently assessed by agarose gel electrophoresis. Briefly, a plasmid, such as pSVβ, is digested with restriction enzymes to yield a variety of fragment sizes. For ease of detection, the fragments may be labeled with ^{32}P either by filling in of the ends with DNA polymerase I or by phosphorylation of the 5'-end with polynucleotide kinase following dephosphorylation by alkaline phosphatase. The plasmid fragments are then incubated with the receptor-binding internalized ligand/nucleic acid binding domain in this case, FGF2-3/poly-L-lysine in a buffered saline solution, such as 20 mM HEPES, pH 7.3, 0.1M NaCl. The reaction mixture is electrophoresed on an agarose gel alongside similarly digested, but nonreacted fragments. If a radioactive label was incorporated, the gel may be dried and autoradiographed. If no radioactive label is present, the gel may be stained with ethidium bromide and the DNA visualized through appropriate red filters after excitation with UV. Binding has occurred if the mobility of the fragments is retarded compared to the control. In the example case, the mobility of the fragments was retarded after binding with the FGF2-3/poly-L-lysine conjugate.

Further testing of the conjugate is performed to show that it binds to the cell surface receptor and is internalized into the cell. It is not necessary that the receptor-binding internalized ligand part of the conjugate retain complete biological activity. For example, HBEGF is mitogenic on certain cell types. As discussed above, this activity may not always be desirable. If this activity is present, a proliferation assay is performed. Likewise, for each desirable activity, an appropriate assay may be performed. However, for application of the subject invention, the only criteria that need be met are receptor binding and internalization.

Receptor binding and internalization may be measured by the following three assays. (1) A competitive inhibition assay of the complex to cells expressing the appropriate receptor demonstrates receptor binding. (2) Receptor binding and internalization may be assayed by measuring β-gal expression (*e.g.*, enzymatic activity) in cells that have been transformed with a complex of a β-gal containing plasmid condensed with a receptor-binding internalized ligand/nucleic acid binding domain. This assay is particularly useful for optimizing conditions to give maximal transformation. Thus, the optimum ratio of receptor-binding internalized ligand/nucleic acid binding domain to nucleic acid and the amount of DNA per cell may readily be

determined by assaying and comparing the enzymatic activity of β-gal. As such, these first two assays are useful for preliminary analysis and failure to show receptor binding or β-gal activity does not per se eliminate a candidate receptor-binding internalized ligand/nucleic acid binding domain conjugate or fusion protein from further analysis. (3) The preferred assay is a cytotoxicity assay performed on cells transformed with a cytocide-encoding agent bound by receptor-binding internalized ligand/nucleic acid binding domain. While, in general, any cytocidal molecule may be used, ribosome-inactivating proteins are preferred and saporin, or another type I ribosome-inactivating protein, is particularly preferred. A statistically significant reduction in cell number demonstrates the ability of the receptor-binding internalized ligand/nucleic acid binding domain conjugate or fusion to deliver nucleic acids into a cell.

C. Other elements

1. Nuclear translocation signals

As used herein, a nuclear translocation or targeting sequence (NTS) is a sequence of amino acids in a protein that are required for translocation of the protein into a cell nucleus. Examples of NTS are set forth in Table 2, below. Comparison with known NTSs, and if necessary testing of candidate sequences, should permit those of skill in the art to readily identify other amino acid sequences that function as NTSs.

As used herein, heterologous NTS refers to an NTS that is different from the NTS that occurs in the wild-type peptide, polypeptide, or protein. For example, the NTS may be derived from another polypeptide, it may be synthesized, or it may be derived from another region in the same polypeptide. A typical consensus NTS sequence contains an amino-terminal proline or glycine followed by at least three basic residues in a array of seven to nine amino acids (*see, e.g.* Dang et al. (1989) *J. Biol. Chem. 264*:18019-18023, Dang et al. (1988) *Mol. Cell. Biol. 8*:4049-4058 and Table 2, which sets forth examples of NTSs and regions of proteins that share homology with known NTSs).

TABLE 2*

Source	Sequence	SEQ ID NO.
SV40 large T	Pro^{126} LysLysArgLysValGlu	67
Polyoma large T	Pro^{279} ProLysLysAlaArgGluVal	68

Human c-Myc	Pro[120]AlaAlaLysArgValLysLeuAsp	69
Adenovirus E1A	Lys[281]ArgProArgPro	70
Yeast mat α2	Lys[3]IleProIleLys	71
c-Erb-A	A. Gly[22]LysArgLysArgLysSer B. Ser[127]LysArgValAlaLysArgLysleu C. Ser[181]HisTrpLysGlnLysArgLysPhe	72 73 74
c-Myb	Pro[521]LeuLeuLysLysIleLysGln	75
p53	Pro[316]GlnProLysLysLysPro	76
Nucleolin	Pro[277]GlyLysArgLysLysGluMetThrLysGlnLysGluValPro	77
HIV Tat	Gly[48]ArgLysLysArgArgGlnArgArgArgAlaPro	78
FGF-1	AsnTyrLysLysProLysLeu	79
FGF-2	HisPheLysAspProLysArg	80
FGF-3	AlaProArgArgArgLysLeu	63
FGF-4	IleLysArgLeuArgArg	-
FGF-5	GlyArgArg	-
FGF-6	IleLysArgGlnArgArg	-
FGF-7	IleArgValArgArg	65
VEGF$_{189}$	LysArgLysArgLysLys (in EXON VI)	66

VEGF206	LysArgLysArgLysLys (in EXON VI)	66
PDGF	ProLysGlyLysHisArgLysPheLysHisThi	

*Superscript indicates position in protein

2. Cytoplasm-translocation signal

Cytoplasm-translocation signal sequence is a sequence of amino acids in a protein that cause retention of proteins in the lumen of the endoplasmic reticulum and/or translocate proteins to the cytosol. The signal sequence in mammalian cells is KDEL (Lys-Asp-Glu-Leu) (Munro and Pelham, *Cell 48*:899-907, 1987). Some modifications of this sequence have been made without loss of activity. For example, the sequences RDEL (Arg-Asp-Glu-Leu) and KEEL (Lys-Glu-Glu-Leu) confer efficient or partial retention, respectively, in plants (Denecke et al., *Embo. J. 11*:2345-2355, 1992).

A cytoplasm-translocation signal sequence may be included in saporin or, for conjugates of HBEGF with a nucleic acid binding domain, the sequence may reside in either part or both. If cleavable linkers are used in the conjugate, the cytoplasm-translocation signal is preferably included in saporin or the nucleic acid binding domain. Additionally, a cytoplasmic-translocation signal sequence may be included in HBEGF, as long as it is placed so as not to interfere with receptor binding.

In addition, or alternatively, membrane-disruptive peptides may be incorporated into complexes of HBEGF-nucleic acid binding domain and cytocide-encoding agent. Adenoviruses are known to enhance disruption of endosomes. Virus-free viral proteins, such as influenza virus hemagglutinin HA-2, may be useful in the present invention. Other proteins may be tested in the assays described herein to find specific endosome disrupting agents that enhance gene delivery. In general, these proteins and peptides are amphipathic (*see*, Wagner et al., *Adv. Drug. Del. Rev. 14*:113-135, 1994).

3. Linkers

A linker is a peptide or other molecule that couples a HBEGF polypeptide to the targeted agent. The linker may be bound via the N- or C-terminus or an internal reside, but, typically within about 20 amino acids of either terminus of a HBEGF and/or targeted agent. The linkers provided herein increase intracellular availability, serum stability, specificity and solubility of the conjugate or provide increased flexibility or relieve steric hindrance in the conjugate. For example,

specificity or intracellular availability of the targeted agent may be conferred by including a linker that is a substrate for certain proteases, such as a protease that is present in only certain subcellular compartments or that is present at higher levels in tumor cells than normal cells.

In order to increase the serum stability, solubility and/or intracellular concentration and to reduce steric hindrance caused by close proximity of HBEGF and the targeted agent, one or more linkers is(are) inserted between the HBEGF protein and the targeted moiety. These linkers include peptide linkers, such as intracellular protease substrates and peptides that increase flexibility or solubility of the linked moieties, and chemical linkers, such as acid labile linkers, ribozyme substrate linkers and others. Peptide linkers may be inserted using heterobiofunctional reagents, described below, or, preferably, are linked to HBEGF by linking DNA encoding the substrate to the DNA encoding the HBEGF protein and expressing the resulting chimera. In instances in which the targeted agent is a protein, such as a RIP, the DNA encoding the linker can be inserted between the DNA encoding the HBEGF protein and the DNA encoding the targeted protein agent.

Chemical linkers may be inserted by covalently coupling the linker to the HBEGF protein and the targeted agent. The heterobifunctional agents, described below, may be used to effect such covalent coupling.

a. Protease substrates

Peptides encoding protease-specific substrates are introduced between the HBEGF protein and the targeted moiety. The peptides may be inserted using heterobiofunctional reagents, described below, or, preferably, are linked to HBEGF by linking DNA encoding the substrate to the DNA encoding the HBEGF protein and expressing the resulting chimera. In instances in which the targeted agent is a protein, such as a RIP, the DNA encoding the linker can be inserted between the DNA encoding the HBEGF protein and the DNA encoding the targeted protein agent. For example, DNA encoding substrates specific for intracellular proteases has been inserted between the DNA encoding the HBEGF protein and a targeted agent, such as saporin.

Any protease specific substrate (*see, e.g.*, O'Hare et al. (1990) *FEBS 273*:200-204; Forsberg et al. (1991) *J. Protein Chem. 10*:517-526; Westby et al. (1992) *Bioconjuugate Chem. 3*:375-381) may be introduced as a linker between the HBEGF polypeptide and linked targeting agent as long as the substrate is cleaved in an intracellular compartment. Preferred substrates include those that are specific for proteases that are expressed at higher levels in tumor cells or that are preferentially expressed in the endosome. The following substrates are among those contemplated

for use in accord with the methods herein: cathepsin B substrate, cathepsin D substrate, trypsin substrate, thrombin substrate, and recombinant subtilisin substrate (XaaAspGluLeu SEQ ID NO. 50, particularly, PheAlaHisTyr, SEQ ID NO. 49).

b. Flexible linkers and linkers that increase the solubility of the conjugates

Flexible linkers and linkers that increase solubility of the conjugates are contemplated for use, either alone or with other linkers, such as the protease specific substrate linkers. Such linkers include, but are not limited to, $(Gly_4Ser)_n$, $(Ser_4Gly)_n$ and $(AlaAlaProAla)_n$ (see, SEQ ID NO. 48) in which n is 1 to 6, preferably 1-4, such as:

(1) Gly_4Ser SEQ ID NO. 40

CCATGGGCGG CGGCGGCTCT GCCATGG

(2) $(Gly_4Ser)_2$ SEQ ID NO. 41

CCATGGGCGG CGGCGGCTCT GGCGGCGGCG GCTCTGCCAT GG

(3) $(Ser_4Gly)_4$ SEQ ID NO. 42

CCATGGCCTC GTCGTCGTCG GGCTCGTCGT CGTCGGGCTC GTCGTCGTCG GGCTCGTCGT CGTCGGGCGC CATGG

(4) $(Ser_4Gly)_2$ SEQ ID NO. 43

CCATGGCCTC GTCGTCGTCG GGCTCGTCGT CGTCGGGCGC CATGG

(5) $(AlaAlaProAla)_n$, where n is 1 to 4, preferably 2 (see, SEQ ID NO.:48)

c. Heterobifunctional cross-linking reagents

Numerous heterobifunctional cross-linking reagents that are used to form covalent bonds between amino groups and thiol groups and to introduce thiol groups into proteins, are known to those of skill in this art (*see, e.g.*, the PIERCE CATALOG, ImmunoTechnology Catalog & Handbook, 1992-1993, which describes the preparation of and use of such reagents and provides a commercial source for such reagents; see, also, *e.g.*, Cumber et al. (1992) *Bioconjugate Chem. 3*:397-401; Thorpe et al. (1987) *Cancer Res. 47*:5924-5931; Gordon et al. (1987) *Proc. Natl. Acad Sci. 84*:308-312; Walden et al. (1986) *J. Mol. Cell Immunol. 2*:191-197; Carlsson et al. (1978) *Biochem. J. 173*: 723-737; Mahan et al. 91987) *Anal. Biochem. 162*:163-170; Wawryznaczak et al. (1992) *Br. J. Cancer 66*:361-366; Fattom et al. (1992) *Infection & Immun. 60*:584-589). These reagents may be used to form covalent bonds between the HBEGF polypeptide(s) with protease substrate peptide linkers and targeted protein agent. These reagents include, but are not limited to: N-succinimidyl-3-(2-pyridyldithio)propionate (SPDP; disulfide linker); sulfosuccinimidyl 6-[3-(2-pyridyldithio)propion-

amido]hexanoate (sulfo-LC-SPDP); succinimidyloxycarbonyl-α-methyl benzyl thiosulfate (SMBT, hindered disulfate linker); succinimidyl 6-[3-(2-pyridyldithio) propionamido]hexanoate (LC-SPDP); sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (sulfo-SMCC); succinimidyl 3-(2-pyridyldithio)butyrate (SPDB; hindered disulfide bond linker); sulfosuccinimidyl 2-(7-azido-4-methylcoumarin-3-acetamide) ethyl-1,3'-dithiopropionate (SAED); sulfo-succinimidyl 7-azido-4-methylcoumarin-3-acetate (SAMCA); sulfosuccinimidyl 6-[alpha-methyl-alpha-(2-pyridyldithio)toluamido]hexanoate (sulfo-LC-SMPT); 1,4-di-[3'-(2'-pyridyldithio)propionamido]butane (DPDPB); 4-succinimidyloxycarbonyl-α-methyl-α-(2-pyridylthio)toluene (SMPT, hindered disulfate linker);sulfosuccinimidyl6[α-methyl-α-(2-pyridyldithio)toluamido]hexanoate (sulfo-LC-SMPT); *m*-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS); m-maleimidobenzoyl-N-hydroxysulfosuccinimide ester (sulfo-MBS); N-succinimidyl(4-iodoacetyl)aminobenzoate (SIAB; thioether linker); sulfosuccinimidyl(4-iodoacetyl)amino benzoate (sulfo-SIAB); succinimidyl4(*p*-maleimidophenyl)butyrate (SMPB); sulfosuccinimidyl4-(*p*-maleimidophenyl)butyrate (sulfo-SMPB); azidobenzoyl hydrazide (ABH). These linkers should be particularly useful when used in combination with peptide linkers, such as those that increase flexibility.

d. Acid cleavable, photocleavable and heat sensitive linkers

Acid cleavable linkers include, but are not limited to, bismaleimideothoxy propane; and adipic acid dihydrazide linkers (*see, e.g.*, Fattom et al. (1992) *Infection & Immun. 60*:584-589) and acid labile transferrin conjugates that contain a sufficient portion of transferrin to permit entry into the intracellular transferrin cycling pathway (*see, e.g.*, Welhöner et al. (1991) *J. Biol. Chem. 266*:4309-4314). Conjugates linked via acid cleavable linkers should be preferentially cleaved in acidic intracellular compartments, such as the endosome.

Photocleavable linkers are linkers that are cleaved upon exposure to light (*see, e.g.*, Goldmacher et al. (1992) *Bioconj. Chem. 3*:104-107, which linkers are herein incorporated by reference), thereby releasing the targeted agent upon exposure to light. Photocleavable linkers that are cleaved upon exposure to light are known (*see, e.g.*, Hazum et al. (1981) in *Pept., Proc. Eur. Pept. Symp., 16th*, Brunfeldt, K (Ed), pp. 105-110, which describes the use of a nitrobenzyl group as a photocleavable protective group for cysteine; Yen et al. (1989) *Makromol. Chem. 190*:69-82, which describes water soluble photocleavable copolymers, including hydroxypropylmethacrylamide copolymer, glycine copolymer, fluorescein copolymer and methylrhodamine copolymer; Goldmacher et al. (1992) *Bioconj. Chem. 3*:104-107, which describes a

cross-linker and reagent that undergoes photolytic degradation upon exposure to near UV light (350 nm); and Senter et al. (1985) *Photochem. Photobiol 42*:231-237, which describes nitrobenzyloxycarbonyl chloride cross linking reagents that produce photocleavable linkages), thereby releasing the targeted agent upon exposure to light. Such linkers would have particular use in treating dermatological or ophthalmic conditions that can be exposed to light using fiber optics. After administration of the conjugate, the eye or skin or other body part can be exposed to light, resulting in release of the targeted moiety from the conjugate. If the toxic moiety is a light activated porphyrin, light-exposure will also activate the porphyrin, thereby causing cell death. Use of photocleavable linkers should permit administration of higher dosages of such conjugates compared to conjugates that release a cytotoxic agent upon internalization. Heat sensitive linkers would also have similar applicability.

D. Expression vectors and host cells for expression of HBEGF or targeted agents

As used herein, vector or plasmid refers to discrete elements that are used to introduce heterologous DNA into cells for either expression of the heterologous DNA or for replication of the cloned heterologous DNA. Selection and use of such vectors and plasmids are well within the level of skill of the art. Expression refers to the process by which nucleic acid is transcribed into mRNA and translated into peptides, polypeptides, or proteins. If the nucleic acid is derived from genomic DNA, expression may, if an appropriate eukaryotic host cell or organism is selected, include splicing of the mRNA.

As used herein, expression vector includes vectors capable of expressing DNA fragments that are in operative linkage with regulatory sequences, such as promoter regions, that are capable of effecting expression of such DNA fragments. Thus, an expression vector refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector that, upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well known to those of skill in the art and include those that are replicable in eukaryotic cells and/or prokaryotic cells and those that remain episomal or may integrate into the host cell genome.

As used herein, operative linkage or operative association of heterologous DNA to regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional and translational stop sites, and other signal sequences, refers to the functional relationship between such DNA and such sequences of nucleotides. For example, operative linkage of heterologous DNA to a promoter

refers to the physical and functional relationship between the DNA and the promoter such that the transcription of such DNA is initiated from the promoter by an RNA polymerase that specifically recognizes, binds to and transcribes the DNA in reading frame. For example, the NTS may be derived from another polypeptide, it may be synthesized, or it may be derived from another region in the same polypeptide.

As used herein, transfection refers to the taking up of DNA or RNA by a host cell. Transformation refers to this process performed in a manner such that the DNA is replicable, either as an extrachromosomal element or as part of the chromosomal DNA of the host. Methods and means for effecting transfection and transformation are well known to those of skill in this art (*see, e.g.*, Wigler et al. (1979) *Proc. Natl. Acad. Sci. USA 76:*1373-1376; Cohen et al. (1972) *Proc. Natl. Acad. Sci. USA 69:*2110).

DNA encoding the selected HBEGF or a portion thereof, HBEGF conjugate or polypeptide targeted agent is inserted into a suitable vector and expressed in a suitable prokaryotic or eukaryotic host. Numerous suitable hosts and vectors are known and available to those of skill in this art and may be purchased commercially or constructed according to published protocols using well known and available starting materials. Suitable eukaryotic host cells include insect cells, yeast cells, and animal cells. Insect cells and bacterial host cells are presently preferred. Suitable prokaryotic host cells include *E. coli*, strains of *Bacillus* and *Streptomyces*.

The plasmids used herein must include a promoter in operable association with the DNA encoding the protein or polypeptide of interest and are designed for expression of proteins in a bacterial host. A promoter region refers to the portion of DNA of a gene that controls transcription of DNA to which it is operatively linked. A portion of the promoter region includes specific sequences of DNA that are sufficient for RNA polymerase recognition, binding and transcription initiation. Promoters, depending upon the nature of the regulation, may be constitutive or regulated. For use herein, inducible promoters are preferred. The promoters are recognized by an RNA polymerase that is expressed by the host. The RNA polymerase may be endogenous to the host or may be introduced by genetic engineering into the host, either as part of the host chromosome or on an episomal element, including a plasmid containing the DNA encoding the saporin-containing polypeptide. Most preferred promoters for use herein are tightly regulated such that, absent induction, the DNA encoding the saporin-containing protein is not expressed. It has been found that tightly regulatable promoters are preferred for expression of saporin. Suitable promoters for expression of proteins and polypeptides herein are widely available and

are well known in the art. For expression of the proteins such promoters are inserted in a plasmid in operative linkage with a control region such as the lac operon. Preferred promoter regions are those that are inducible and functional in *E. coli* or early genes in vectors of viral origin. Examples of suitable inducible promoters and promoter regions include, but are not limited to: the *E. coli* lac operator responsive to isopropyl β -D-thiogalactopyranoside (IPTG; see, et al. Nakamura et al. (1979) *Cell 18:*1109-1117); the metallothionein promoter metal-regulatory-elements responsive to heavy-metal (*e.g.*, zinc) induction (*see, e.g.*, U.S. Patent No. 4,870,009 to Evans *et al.*); the phage T7lac promoter responsive to IPTG (*see, e.g.*, U.S. Patent No. 4,952,496; and Studier et al. (1990) *Meth. Enzymol. 185:*60-89) and the tac promoter. Other promoters include, but are not limited to, the T7 phage promoter and other T7-like phage promoters, such as the T3, T5 and SP6 promoters, the trp, lpp, and lac promoters, such as the lacUV5, from *E. coli*; the P10 or polyhedrin gene promoter of baculovirus/insect cell expression systems (*see, e.g.*, U.S. Patent Nos. 5,243,041, 5,242,687, 5,266,317, 4,745,051, and 5,169,784) and inducible promoters from other eukaryotic expression systems.

The DNA construct is introduced into a plasmid suitable for expression in the selected host. The sequences of nucleotides in the plasmids that are regulatory regions, such as promoters and operators, are operationally associated with one another for transcription. The sequence of nucleotides encoding the HBEGF, HBEGF chimera or cytotoxic agent may also include DNA encoding a secretion signal, whereby the resulting peptide is a precursor protein. Secretion signals suitable for use are widely available and are well known in the art. Secretion signal refers to a peptide region within the precursor protein that directs secretion of the precursor protein from the cytoplasm of the host into the periplasmic space or into the extracellular growth medium. Such signals may be either at the amino terminus or carboxyl terminus of the precursor protein. The preferred secretion signal is linked to the amino terminus and may be heterologous to the protein to which it is linked. Prokaryotic and eukaryotic secretion signals functional in *E. coli*, may be employed. The presently preferred secretion signals include, but are not limited to, those encoded by the following *E. coli* genes: ompA, ompT, ompF, ompC, beta-lactamase, pelB and bacterial alkaline phosphatase, and the like (von Heijne (1985) *J. Mol. Biol. 184:*99-105). In addition, the bacterial pelB gene secretion signal (Lei et al. (1987) *J. Bacteriol. 169:*4379), the phoA secretion signal, and the cek2 secretion signal, functional in insect cells, may be employed. The most preferred secretion signal for bacterial expression is the *E. coli* ompA secretion signal. For eukaryotic expression systems, particularly insect cell systems, the signals from secreted proteins, such as insulin, growth hormone, mellitin,

and mammalian alkaline phosphatase are of interest herein. Other prokaryotic and eukaryotic secretion signals known to those of skill in the art may also be employed (*see, e.g.,* von Heijne (1985) *J. Mol. Biol. 184:*99-105). Using the methods described herein, one of skill in the art can substitute secretion signals that are functional in either yeast, insect or mammalian cells to secrete the heterologous protein from those cells. The resulting processed protein may be recovered from the periplasmic space or the fermentation medium or growth medium.

The plasmids may also include a selectable marker gene or genes that are functional in the host. A selectable marker gene includes any gene that confers a phenotype on bacteria that allows transformed bacterial cells to be identified and selectively grown from among a vast majority of untransformed cells. Suitable selectable marker genes for bacterial hosts, for example, include the ampicillin resistance gene (Amp^r), tetracycline resistance gene (Tc^r) and the kanamycin resistance gene (Kan^r). The kanamycin resistance gene is presently preferred.

Particularly preferred plasmids for transformation of *E. coli* cells include the pET expression vectors (see, U.S patent 4,952,496; available from Novagen, Madison, WI; see, also literature published by Novagen describing the system). Such plasmids include pET 11a, which contains the T7lac promoter, T7 terminator, the inducible *E. coli* lac operator, and the lac repressor gene; pET 12a-c, which contains the T7 promoter, T7 terminator, and the *E. coli* ompT secretion signal; and pET 15b (Novagen, Madison, WI), which contains a His-Tag™ leader sequence (Seq. ID NO. 23) for use in purification with a His column and a thrombin cleavage site that permits cleavage following purification over the column; the T7-lac promoter region and the T7 terminator.

Other preferred plasmids include the pKK plasmids, particularly pKK 223-3, which contains the TAC promoter, (available from Pharmacia; *see also*, Brosius et al. (1984) *Proc.. Natl. Acad. Sci. 81:*6929; Ausubel et al., Current Protocols in Molecular Biology; U.S. Patent Nos. 5,122,463, 5,173,403, 5,187,153, 5,204,254, 5,212,058, 5,212,286, 5,215,907, 5,220,013, 5.223,483, and 5,229,279). Plasmid pKK has been modified by insertion of a kanamycin resistance cassette with *Eco*RI sticky ends (purchased from Pharmacia; obtained from pUC4K, *see, e.g.,* Vieira et al. (1982) *Gene 19:*259-268; and U.S. Patent No. 4,719,179) into the ampicillin resistance marker gene.

Other preferred vectors include the pP_L-lambda inducible expression vector, pTrc99A, and the *tac* promoter vector pDR450 (*see, e.g.,* U.S. Patent Nos. 5,281,525, 5,262,309, 5,240,831, 5,231,008, 5,227,469, 5,227,293, ; available from

Pharmacia P.L. Biochemicals, see; also Mott, et al. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82:*88; and De Boer et al. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80:*21); and baculovirus vectors, such as a pBlueBac vector (also called pJVETL and derivatives thereof; *see, e.g.,* U.S. Patent Nos. 5,278,050, 5,244,805, 5,243,041, 5,242,687, 5,266,317, 4,745,051, and 5,169,784), including pBlueBac III.

Other plasmids include the pIN-IIIompA plasmids (see, U.S. Patent No. 4,575,013 to Inouye; see, also, Duffaud et al. (1987) *Meth. Enz. 153:*492-507), such as pIN-IIIompA2 . The pIN-IIIompA plasmids include an insertion site for heterologous DNA linked in transcriptional reading frame with functional fragments derived from the lipoprotein gene of *E. coli.* The plasmids also include a DNA fragment coding for the signal peptide of the ompA protein of *E. coli*, positioned such that the desired polypeptide is expressed with the ompA signal peptide at its amino terminus, thereby allowing efficient secretion across the cytoplasmic membrane. The plasmids further include DNA encoding a specific segment of the *E. coli* lac promoter-operator, which is positioned in the proper orientation for transcriptional expression of the desired polypeptide, as well as a separate functional *E. coli* lacI gene encoding the associated repressor molecule that, in the absence of lac operon inducer, interacts with the lac promoter-operator to prevent transcription therefrom. Expression of the desired polypeptide is under the control of the lipoprotein (lpp) promoter and the lac promoter-operator, although transcription from either promoter is normally blocked by the repressor molecule. The repressor is selectively inactivated by means of an inducer molecule thereby inducing transcriptional expression of the desired polypeptide from both promoters.

The repressor protein may be encoded by the plasmid containing the construct or a second plasmid that contains a gene encoding for a repressor-protein. The repressor-protein is capable of repressing the transcription of a promoter that contains sequences of nucleotides to which the repressor-protein binds. The promoter can be derepressed by altering the physiological conditions of the cell. The alteration can be accomplished by the addition to the growth medium of a molecule that inhibits, for example, the ability to interact with the operator or with regulatory proteins or other regions of the DNA or by altering the temperature of the growth media. Preferred repressor-proteins include, but are not limited to the *E. coli* lacI repressor responsive to IPTG induction, the temperature sensitive cI857 repressor, and the like. The *E. coli* lacI repressor is preferred.

In certain preferred embodiments, the constructs also include a transcription terminator sequence. A transcription terminator region has either (a) a

subsegment that encodes a polyadenylation signal and polyadenylation site in the transcript, and/or (b) a subsegment that provides a transcription termination signal that terminates transcription by the polymerase that recognizes the selected promoter. The entire transcription terminator may be obtained from a protein-encoding gene, which may be the same or different from the gene, which is the source of the promoter. Preferred transcription terminator regions are those that are functional in *E. coli*. Transcription terminators are optional components of the expression systems herein, but are employed in preferred embodiments. The promoter regions and transcription terminators are each independently selected from the same or different genes. In some embodiments, the DNA fragment is replicated in bacterial cells, preferably in *E. coli*. The DNA fragment also typically includes a bacterial origin of replication, to ensure the maintenance of the DNA fragment from generation to generation of the bacteria. In this way, large quantities of the DNA fragment can be produced by replication in bacteria. Preferred bacterial origins of replication include, but are not limited to, the f1-ori and col E1 origins of replication.

Preferred bacterial hosts contain chromosomal copies of DNA encoding T7 RNA polymerase operably linked to an inducible promoter, such as the lacUV promoter (see, U.S. Patent No. 4,952,496). Such hosts include, but are not limited to, lysogenic *E. coli* strains HMS174(DE3)pLysS, BL21(DE3)pLysS, HMS174(DE3) and BL21(DE3). Strain BL21(DE3) is preferred. The pLys strains provide low levels of T7 lysozyme, a natural inhibitor of T7 RNA polymerase. Preferred eukaryotic hosts are the insect cells *Spodoptera frugiperda* (sf9 cells; *see, e.g.*, Luckow et al. (1988) *Bio/technology 6:*47-55 and U.S. Patent No. 4,745,051).

For insect hosts, baculovirus vectors, such as a pBlueBac vector (also called pJVETL and derivatives thereof), particularly pBlueBac III, (*see, e.g.*, U.S. Patent Nos. 5,278,050, 5,244,805, 5,243,041, 5,242,687, 5,266,317, 4,745,051, and 5,169,784; available from INVITROGEN, San Diego) may also be used for expression of the polypeptides. The pBlueBacIII vector is a dual promoter vector and provides for the selection of recombinants by blue/white screening as this plasmid contains the β-galactosidase gene (lacZ) under the control of the insect recognizable ETL promoter and is inducible with IPTG. A DNA construct is introduced into a baculovirus vector pBluebac III (INVITROGEN, San Diego, CA) and then co-transfected with wild type virus into insect cells *Spodoptera frugiperda* (sf9 cells; *see, e.g.*, Luckow et al. (1988) *Bio/technology 6:*47-55 and U.S. Patent No. 4,745,051).

Other baculovirus vectors, such as pPbac and pMbac (available from Stratagene, San Diego, CA, see, also Lernhardt et al. (1993) *Strategies 6:*20-21, and the

Stratagene Catalog page 218), which contain the human alkaline phosphatase (*see, e.g.*, Bailey et al. (1989) *Proc. Natl. Acad. Sci. U. S. A. 86:*22-26) and melittin (*see, e.g.*, Tessier et al. (1991) *Gene 98:*177-183) secretory signals inserted into the BamHI and NdeI sites, respectively of pJVP10Z (*see, e.g.*, Kawamoto et al. (1991) *Biochem. Biophys. Res. Commun. 181:*756-63, Ueda et al.(1994) *Gene 140:*267-272, are also suitable for use herein, particularly if secretion is desired. Insertion of genes into the *SmaI/Bam*HI sites of these vectors results in fusion proteins that are directed into the insect cell secretory pathway, which processes the pro-polypeptide so that mature peptide or fusion protein is secreted into the growth medium. Other heterologous signal sequences, such as the insulin signal sequence (*see, e.g.*, U.S. Patent No. 4,431,746 for DNA encoding the signal sequence), the growth hormone signal sequence, mammalian alkaline phosphatase, the mellitin signal sequence and others that are processed by insect cells are used.

DNA encoding full-length HBEGF, HBEGF-SAP, SAP-HBEGF with and without linkers, and other such constructs, has been introduced into the pET vectors pET 11a (Novagen, Madison, WI). DNA encoding SAP has also been introduced in pET 15b (Novagen, Madison, WI).

Some of the constructs provided herein have also been inserted into the baculovirus vector sold commercially under the name pbluebacIII (Invitrogen, San Diego CA; see the Invitrogen Catalog; see, Vialard et al. (1990) *J. Virol. 64:*37; *see also*, U.S. Patent No. 5,270,458; U.S. Patent No. 5,243,041; and published International PCT Application WO 93/10139, which is based on U.S. patent application Serial No. 07/792,600. The pBlueBacIII vector is a dual promoter vector and provides for the selection of recombinants by blue/white screening as this plasmid contains the β-galactosidase gene (lacZ) under the control of the insect recognizable ETL promoter and is inducible with IPTG. The HBEGF construct or other construct is inserted into this vector under control of the polyhedrin promoter. The DNA is then cotransfected, such as by $CaPO_4$ transfection or liposomes, into *Spodoptera frugiperda* cells (sf9 cells) with wild type baculovirus and grown in tissue culture flasks or in suspension cultures. Blue occlusion minus viral plaques are selected and plaque purified and screened for the presence of HBEGF-encoding DNA by any standard methodology, such as western blots using HBEGF anti-sera or Southern blots using an appropriate HBEGF probe. DNA encoding an HBEGF with and without linkers is introduced into a Bluebac vector for expression in baculovirus. Details are set forth in the Examples.

E. Methods of preparation of HBEGF-targeted agent conjugates

Cytotoxic conjugates with linked targeted agents can be prepared either by chemical conjugation, recombinant DNA technology, or combinations of recombinant expression and chemical conjugation. The methods herein are exemplified with particular reference to HBEGF and saporin. It is understood, however, that the same methods may be used to prepare and use conjugates of any HBEGF polypeptide with any cytotoxic agent, such as a RIP, a nucleic acid or any other targeted agent either directly or via linkers as described herein. The growth factor and targeted agent may be linked in any orientation and more than one growth factor and/or targeted agent may be present in a conjugate.

Conjugates that contain one or more HBEGF polypeptides linked, either directly or via a linker, to one or more targeted agents are provided. The presently preferred HBEGF polypeptides are those having sequences set forth in SEQ ID NOs. 1-5. Human HBEGF is particularly preferred.

The conjugates provided herein contain the following components: $(HBEGF)_n$, $(L)_q$, and $(\text{targeted agent})_m$ in which: at least one HBEGF moiety is linked with or without a linker (L) to at least one targeted agent, n is 1 or more, typically is between 2 and 6, generally 1 or 2; q is 0 or more as long as the resulting conjugate binds to the targeted receptor, is internalized and delivers the targeted agent, q is generally 1 to 4; m is 1 or more, generally 1 or 2; L refers to a linker, and the targeted agent is any agent, such as a cytotoxic agent or a nucleic acid, or a drug, such as methotrexate, intended for internalization by a cell that expresses a receptor to which HBEGF binds and upon binding is internalized.

It is understood that the HBEGF and targeted agent (or linker and targeted agent) may be linked in any order and through any appropriate linkage, as long as the resulting conjugate binds to a receptor to which HBEGF binds and internalizes the targeted agent(s) in cells bearing the receptor.

For example, the HBEGF polypeptide may be linked to the targeted agent or linker at or near its N-terminus or at or near its C-terminus. The HBEGF may be linked to a second HBEGF polypeptide, which may be the same or a different HBEGF polypeptide; and one or more targeted agents may be linked to the HBEGF or may be linked to each other. The linkage may be at any locus, although the C- or N-terminus region of HBEGF (within about 20, preferably 10, amino acids from the terminus) is preferred. If more than one targeted agent is included, the second agent may be the same or different from the first agent.

In some embodiments, the conjugates provided herein may be represented by the formulae (I):

$(HBEGF_n\text{-}(L)_q\text{-targeted agent}_m)_p$

in which HBEGF refers to a polypeptide that is reactive with a HBEGF receptor (also referred to herein as a HBEGF polypeptide), such as HBEGF, L refers to a linker, which may be present or absent, q is 0 or more as long as the resulting conjugate binds to a targeted receptor and the targeted agent is internalized, m, n and p are, independently, 1 or more, and generally less than or equal to 4, and preferably 1 or 2, and the targeted agent is any agent, such as a cytotoxic agent or a nucleic acid, or a drug, such as methotrexate, intended for internalization by a cell that expresses a HBEGF receptor; and the HBEGF may be linked to the linker or targeted agent via its N-terminus or C-terminus or any other locus in polypeptide, such as derivatized cys residues. When p is 2, the conjugates are preferably linked via cysteine residues present or introduced into HBEGF.

Conjugates of the formula (II): $((\text{targeted agent})_m\text{-}(L)_q\text{-}(HBEGF)_n)_p$, in which m, n, p and 1 are as defined above, are also provided. These conjugates are prepared by chemical conjugation or by preparing fusion proteins from DNA constructs that encode one or two HBEGF moieties.

In addition, conjugates in which non-essential cysteines in the HBEGF polypeptides and/or targeted agent, if the agent is a polypeptide, are deleted or replaced with Ser or other conservative substitution are provided. Compositions of such conjugates should exhibit reduced aggregation compared to conjugates that contain non-essential cysteines. Non-essential cysteines may be identified empirically.

The linker is selected to increase the specificity, toxicity, solubility, serum stability, and/or intracellular availability of the targeted moiety. More preferred linkers are those that can be incorporated in fusion proteins and expressed in a host cell, such as *E. coli*. Such linkers include: enzyme substrates, such as cathepsin B substrate, cathepsin D substrate, trypsin substrate, thrombin substrate, subtilisin substrate, factor Xa substrate, and enterokinase substrate; linkers that increase solubility, flexibility, and/or intracellular cleavability, such as $(gly_mser)_n$ and $(ser_mgly)_n$, in which n is 1 to 6, preferably 1 to 4, most preferably 1, and m is 1 to 6, preferably 1 to 4, more preferably 4. Preferred among such linkers, are those, such as cathepsin D substrate, that are preferentially cleaved in the endosome or cytoplasm following internalization of the conjugate linker; other such linkers, such as $(gly_mser)_n$ and $(ser_mgly)_n$, also increase the flexibility, serum stability and/or solubility of the conjugate or the availability of the region joining the HBEGF and targeted agent for cleavage. In some

embodiments, several linkers that are the same or different may be included in order to take advantage of desired properties of each linker.

Other linkers are suitable for incorporation into chemically produced conjugates. Linkers that are suitable for chemically linking conjugates include disulfide bonds, thioether bonds, hindered disulfide bonds, and covalent bonds between free reactive groups, such as amine and thiol groups. These bonds are produced using heterobifunctional reagents to produce reactive thiol groups on one or both of the polypeptides and then reacting the thiol groups on one polypeptide with reactive thiol groups or amine groups to which reactive maleimido groups or thiol groups can be attached on the other. Other linkers include acid cleavable linkers, such as bismaleimideothoxy propane, acid labile-transferrin conjugates and adipic acid diihydrazide, that would be cleaved in more acidic intracellular compartments and cross linkers that are cleaved upon exposure to UV or visible light and linkers.

The targeted agents or moieties include any molecule that, when internalized, alter the metabolism or gene expression in the cell . Such agents include cytotoxic agents, such as ribosome inactivating proteins DNA encoding cytotoxic agents, and antisense nucleic acids, that result in inhibition of growth or cell death. Other such agents also include antisense RNA, DNA, and truncated proteins that alter gene expression via interactions with the DNA, or co-suppression or other mechanism. The conjugates herein may also be used to deliver DNA and thereby serve as agents for gene therapy or to deliver agents that, upon, transcription and/or translation thereof, result in cell death. Cytotoxic agents include, but are not limited to, ribosome inactivating proteins, inhibitors of DNA, RNA and/or protein synthesis, including antisense nucleic acids, and other metabolic inhibitors. In certain embodiments, the cytotoxic agent is a ribosome-inactivating protein (RIP), such as, for example, saporin, although other cytotoxic agents can also be advantageously used.

The targeted agents may also be modified to render them more suitable for conjugation with the linker and/or a HBEGF protein or to increase their intracellular activity. Such modifications include, but are not limited to, the introduction of a Cys residue at or near the N-terminus or C-terminus, derivatization to introduce reactive groups, such as thiol groups, and addition of sorting signals, such as $(XaaAspGluLeu)_n$ (SEQ ID NO. 50), where Xaa is Lys or Arg, preferably Lys, and n is 1 to 6, preferably 1-3, at, preferably, the carboxy-terminus (*see, e.g.*, Seetharam et al. (1991) *J. Biol. Chem. 266*:17376-17381; and Buchner et al. (1992) *Anal. Biochem. 205*:263-270), that direct the targeted agent to the endoplasmic reticulum or the addition of a cytoplasmic sorting sequence, such as KDEL (see discussion herein).

Conjugates that contain a plurality of HBEGF polypeptides linked to the cytotoxic agent are also provided. These conjugates that contain several, typically two to about six, monomers can be produced by linking multiple copies of DNA encoding the HBEGF fusion protein under the transcriptional control of a single promoter region. In addition conjugates that contain, more than one targeted agent per HBEGF, such as SAP-HBEGF-SAP, linked with or without linkers are contemplated herein.

1. Chemical conjugation methods

a. Preparation of HBEGF polypeptides for chemical conjugation

HBEGF may be isolated from a suitable source or may be produced using recombinant DNA methodology, discussed below.

As used herein, "substantially pure" means sufficiently homogeneous to appear free of readily detectable impurities as determined by standard methods of analysis, such as thin layer chromatography (TLC), gel electrophoresis, high performance liquid chromatography (HPLC), used by those of skill in the art to assess such purity, or sufficiently pure such that further purification would not detectably alter the physical and chemical properties, such as enzymatic and biological activities, of the substance. Methods for purification of the compounds to produce substantially chemically pure compounds are known to those of skill in the art. A substantially chemically pure compound may, however, be a mixture of stereoisomers. In such instances, further purification might increase the specific activity of the compound.

To effect chemical conjugation herein, the HBEGF polypeptide is conjugated generally via a reactive amine group or thiol group to the targeted agent or to a linker, which has been or is subsequently linked to the targeted agent. The HBEGF polypeptide is conjugated either via its N-terminus, C-terminus, or elsewhere in the polypeptide. In preferred embodiments, the HBEGF polypeptide is conjugated via a reactive cysteine residue to the linker or to the targeted agent. The HBEGF can also be modified by addition of a cysteine residue, either by replacing a residue or by inserting the cysteine, at or near the amino or carboxyl terminus, within about 20, preferably 10 residues from either end, and preferably at or near the amino terminus.

In order to reduce the heterogeneity of preparations, the HBEGF polypeptide can be modified by mutagenesis to replace reactive cysteines, leaving, preferably, only one available cysteine for reaction. The HBEGF polypeptide is modified by deleting or replacing a site(s) on the HBEGF that causes the heterogeneity. Such sites are typically cysteine residues that, upon folding of the protein, remain available for interaction with other cysteines or for interaction with more than one

cytotoxic molecule per molecule of HBEGF polypeptide. Thus, such cysteine residues do not include any cysteine residues that are required for proper folding of the HBEGF polypeptide or for retention of the ability to bind to a HBEGF receptor and internalize. For chemical conjugation, one cysteine residue that, in physiological conditions, is available for interaction, is not replaced because it is used as the site for linking the cytotoxic moiety. The resulting modified HBEGF is conjugated with a single species of cytotoxic conjugate. Alternatively, the contribution of each cysteine to the ability to bind to HBEGF receptors may be determined empirically as described herein.

b. Preparation of targeted proteins for chemical conjugation

If the targeted agent is a polypeptide it may be directly linked to the HBEGF or HBEGF with linker or to a linker by reaction of a reactive group in the polypeptide. It is desirable, however, that the agent may react at only a single locus, so that the resulting preparation of conjugates is homogeneous. Thus, if necessary, the targeted agent can be derivatized and then a single species isolated, or can be modified so that it only has one reactive group, such as a cysteine, for a particular set of conditions and reagents. For example, saporin has been derivatized and a single species isolated. Saporin has also been modified by introduction of a single cysteine residue.

For chemical conjugation, the SAP may be derivatized or modified such that it includes a cysteine residue for conjugation to the HBEGF protein.

Saporin for chemical conjugation may be produced by isolating the protein from the leaves or seeds of *Saponaria officinalis* or using recombinant methods and the DNA provided herein or known to those of skill in the art or obtained by screening appropriate libraries (*see, e.g.*, International PCT Application WO 93/25688, which describes the isolation of saporin, plasmids containing DNA encoding saporin, expression of saporin and isolation of purified saporin). Some DNA encoding saporin may also include an N-terminal extension sequence linked to the amino terminus of the saporin that encodes a linker so that, if desired, the SAP and linker can be expressed as a fusion protein. The sequence of DNA encoding saporin is set forth in SEQ ID Nos. 8-12.

The DNA molecules provided herein encode saporin that has substantially the same amino acid sequence and ribosome-inactivating activity as that of saporin-6 (SO-6), including any of four isoforms, which have heterogeneity at amino acid positions 48 and 91 (*see, e.g.*, Maras et al. (1990) *Biochem. Internat. 21*:631-638 and Barra et al. (1991) *Biotechnol. Appl. Biochem. 13*:48-53 and SEQ ID NOs. 8-12). Other suitable saporin polypeptides include other members of the multi-gene family coding for isoforms of saporin-type RIPs including SO-1 and SO-3 (Fordham-Skelton

et al. (1990) *Mol. Gen. Genet. 221*:134-138), SO-2 (*see, e.g.*, U.S. Application Serial No. 07/885,242, which corresponds to GB 2,216,891; see, also, Fordham-Skelton et al. (1991) *Mol. Gen. Genet. 229*:460-466), SO-4 (*see, e.g.*, GB 2,194,241 B; see, also, Lappi et al. (1985 *Biochem. Biophys. Res. Commun. 129*:934-942) and SO-5 (*see, e.g.*, GB 2,194,241 B; see, also, Montecucchi et al. (1989) *Int. J. Peptide Protein Res., 33*:263-267; and Ferreras et al. (1993) *Biophys. Biochem. Acta 1216*:31-42). SO-4, which includes the N-terminal 40 amino acids set forth in SEQ ID NO. 13, is isolated from the leaves of *Saponaria officinalis* by extraction with 0.1 M phosphate buffer at pH 7, followed by dialysis of the supernatant against sodium borate buffer, pH 9, and selective elution from a negatively charged ion exchange resin, such as Mono S (Pharmacia Fine Chemicals, Sweden) using a gradient of 1 to 0.3 M NaCl and is the first eluting chromatographic fraction that has SAP activity. The second eluting fraction is SO-5.

Because more than one amino group on SAP may react with the succinimidyl moiety, it is possible that more than one amino group on the surface of the protein is reactive. This creates the potential for heterogeneity even if mono-derivatized SAP is used. This source of heterogeneity has been solved by the conjugating modified SAP expressed in *E. coli* that has an additional cysteine inserted in the coding sequence, preferably within 10 or 20 amino acids of either the C-terminus or N-terminus.

As discussed above, muteins of saporin that contain a Cys at or near the amino or carboxyl terminus can be prepared. Thus, instead of derivatizing saporin to introduce a sulfhydryl, the saporin can be modified by the introduction of a cysteine residue into the SAP such that the resulting modified saporin protein reacts with a HBEGF monomer or a linker (and then to a HBEGF monomer) to produce a conjugate.

Preferred loci for introduction of a cysteine residue include the N-terminus region, preferably within about one to twenty residues, more preferably one to about ten residues, from the N-terminus of the cytotoxic agent, such as SAP. For expression of SAP in the bacterial host systems herein, it is also desirable to add DNA encoding a methionine linked to the DNA encoding the N-terminus of the saporin protein. DNA encoding SAP has been modified by inserting a DNA encoding Met-Cys (ATG TGT or ATG TGC) at the N-terminus immediately adjacent to the codon for first residue of the mature protein.

Muteins in which a cysteine residue has been added at the N-terminus and muteins in which the amino acid at position 4 or 10 has been replaced with cysteine have been prepared by modifying the DNA encoding saporin (see, Example 3). The modified DNA may be expressed and the resulting saporin protein purified, as

described herein for expression and purification of the resulting SAP. The modified saporin can then be reacted with an HBEGF, to form disulfide linkages between the HBEGF and the cysteine residue on the modified SAP.

Typically, SAP is derivatized by reaction with SPDP. This results in a heterogeneous population. For example, SAP that is derivatized by SPDP to a level of 0.9 moles pyridine-disulfide per mole of SAP includes a population of non-derivatized, mono-derivatized and di-derivatized SAP. Methods for isolation of mono-derivatized saporin are described, for example, in Lappi et al. (1993) *Anal. Biochem. 212*:446-451, copending U.S. Application Serial No. 08/099,924). The methods rely on the charge differences among the three species of SAP that are produced upon reaction of one or more lysines in saporin with SPDP. The mono-derivatized saporin species is purified by Mono-S cation exchange chromatography and pooling of the fractions that contain the monoderivatized species. Briefly, as described in the copending application, the initial eluting peak is composed of SAP that is approximately di-derivatized; the second peak is mono-derivatized and the third peak shows no derivatization. The di-derivatized material accounts for 20% of the three peaks; the second accounts for 48% and the third peak contains 32%. Fractions that have a ratio of SPDP to SAP greater than 0.85 but less than 1.05 are pooled, dialyzed against an appropriate buffer, such as 0.1 M sodium chloride, 0.1 M sodium phosphate, pH 7.5, used for coupling to a linker, to a HBEGF or to HBEGF with linker.

The resulting preparation, although more uniform, is heterogeneous because native saporin as purified from the seed is a mixture of four isoforms, as judged by protein sequencing (*see, e.g.*, copending published International PCT Application WO 93/25688 (Serial No. PCT/US93/05702), which is a continuation-in-part of copending United States Application Serial No. 07/901,718; *see also*, Montecucchi et al. (1989) *Int. J. Pept. Prot. Res. 33*:263-267; Maras et al. (1990) *Biochem. Internat. 21*:631-638; and Barra et al. (1991) *Biotechnol. Appl. Biochem. 13*:48-53). This creates some heterogeneity in the conjugates, since the reaction with SPDP probably occurs equally each isoform. This source of heterogeneity can be removed by using saporin expressed in *E. coli*.

c. Chemical conjugation of an HBEGF polypeptide to linkers and targeted agents

The HBEGF polypeptides are preferably linked via non-essential cysteine residues to the linkers or to the targeted agent. HBEGF that has been modified by introduction of a cys residue at or near one terminus; the N-terminus is preferred; is

used in chemical conjugation (see Examples for preparation of such modified HBEGF). Methods for coupling proteins to the linkers, such as the heterobifunctional agents, or to nucleic acids, or to proteins are known to those of skill in the art and are also described herein.

Methods for chemical conjugation of proteins are known to those of skill in the art. The preferred methods for chemical conjugation depend on the selected components, but preferably rely on disulfide bond formation.

2. Fusion protein of an HBEGF polypeptide and targeted agent

Expression of DNA encoding a fusion of a HBEGF polypeptide linked to the targeted agent results in a more homogeneous preparation of cytotoxic conjugates and is suitable for use, when the selected targeting agent and linker are polypeptides. Aggregate formation can be reduced in preparations containing the fusion proteins by modifying the HBEGF, such as by removal of nonessential cysteines in the heparin-binding domain (amino acids 1-45) and/or the targeted agent to prevent interactions between each conjugate, such as via unreacted cysteines.

a. Expression of HBEGF

DNA encoding the HBEGF polypeptide may be isolated, synthesized or obtained from commercial sources or prepared as described herein in Example 4 and in International Application WO/92/06705 (and the corresponding U.S. patent application serial No. 07/598,082), and Abraham et al. (1993) *Biochem. Biophy. Res. Comm. 190:*125-133. Expression of recombinant HBEGF polypeptides may be performed as described herein; and DNA encoding HBEGF polypeptides may be used as the starting materials for the methods herein.

DNA encoding HBEGF polypeptides and/or the amino acid sequences of HBEGFs are known to those of skill in this art (*see, e.g.,* SEQ ID NOs. 1-5). DNA may be prepared synthetically based on the amino acid sequence or known DNA sequence of an HBEGF or may be isolated using methods known to those of skill in the art or obtained from commercial or other sources known to those of skill in this art. For example, suitable methods are described in Example 4 for amplifying HBEGF encoding cDNA from well known plasmids (*e.g.,* pMTN-HBEGF, ATCC #40900 and pAX-HBEGF, ATCC #40899) containing HBEGF encoding cDNA.

Such DNA may then be mutagenized using standard methodologies to delete or delete and replace any cysteine residues, as described herein, that are responsible for aggregate formation. If necessary, the identity of cysteine residues that contribute to aggregate formation may be determined empirically, by deleting and/or deleting and replacing a cysteine residue and ascertaining whether the resulting HBEGF

with the deleted cysteine forms aggregates in solutions containing physiologically acceptable buffers and salts. Loci for insertion of cysteine residues may also be determined empirically. Generally, regions at or near (within 20, preferably 10 amino acids) the C- or, preferably, the N-terminus are preferred.

The DNA construct encoding the conjugate can be inserted into a plasmid and expressed in a selected host, as described above, to produce a recombinant HBEGF-toxin conjugate. Multiple copies of the modified HBEGF-cytotoxic agent chimera or modified HBEGF-cytotoxic agent chimera can be inserted into a single plasmid in operative linkage with one promoter. When expressed, the resulting protein will be an HBEGF-cytotoxic agent multimer. Typically two to six copies of the chimera are inserted, preferably in a head to tail fashion, into one plasmid.

b. Preparation of muteins for recombinant production of the conjugates

For recombinant expression using the methods described herein, up to all cysteines in the HBEGF polypeptide that are not required for biological activity can be deleted or replaced. Alternatively, for use in the chemical conjugation methods herein, all except for one of these cysteines, which will be used for chemical conjugation to the cytotoxic agent, can be deleted or replaced. Each of the HBEGF polypeptides described herein have six cysteine residues. Each of the six cysteines may independently be replaced and the resulting mutein tested for the ability to bind to HBEGF receptors and to be internalized. Alternatively, the resulting mutein-encoding DNA is used as part of a construct containing DNA encoding the cytotoxic agent linked to the HBEGF-encoding DNA. The construct is expressed in a suitable host cell and the resulting protein tested for the ability to bind to HBEGF receptors and internalize the cytotoxic agent. As long as this ability is retained the mutein is suitable for use herein.

c. DNA constructs and expression of the constructs

DNA encoding HBEGF conjugates is expressed in any suitable host, particularly bacterial and insect hosts. Methods and plasmids for such expression are set forth in the Examples (see, also TABLE 3). Using the methods and materials described above and in the Examples numerous chemical conjugates and fusion proteins have been synthesized. These include those set forth in TABLE 3 below.

Particular details of the syntheses of the conjugates and DNA constructs are set forth in the Examples. The constructs have been prepared and have been or can be inserted into plasmids including pET 11 (with and without the T7 transcription terminator), pET 12 and pET 15 (INVITROGEN, San Diego), λpPL and pKK223-3 (PHARMACIA, P.L.) and derivatives of pKK223-3. The resulting plasmids have been

and can be transformed into bacterial hosts including BL21(DE3), BL21(DE3)+pLYS S, HMS174(DE3), HMS174(DE3)+pLYS S (Novagen, Madison, WI) and N4830(cI857) (see, Gottesman et al. (1980) *J. Mol. Biol. 140:*57-75, commercially available from PL Biochemicals, Inc, also, *see, e.g.,* U.S. Patent Nos. 5,266,465, 5,260,223, 5,256,769, 5,256,769, 5,252,725, 5,250,296, 5,244,797, 5,236,828, 5,234,829, 5,229,273, 4,798,886, 4,849,350, 4,820,631 and 4,780,313). N4830 harbors a heavily deleted phage lambda prophage carrying the mutant c1857 temperature sensitive repressor and an active N gene. The constructs have also been introduced into a baculovirus vector sold commercially under the name pBLUEBACIII (INVITROGEN, San Diego CA; see the INVITROGEN CATALOG; see, also, Vialard et al. (1990) *J. Virol. 64:*37; U.S. Patent No. 5,270,458; U.S. Patent No. 5,243,041; and published International PCT Application WO 93/10139, which is based on U.S. patent application Serial No. 07/792,600. The pBlueBacIII vector is a dual promoter vector and provides for the selection of recombinants by blue/white screening as this plasmid contains the β-galactosidase gene (lacZ) under the control of the insect recognizable ETL promoter and is inducible with IPTG. The baculovirus vector is then co-transfected with wild type virus into insect host cells *Spodoptera frugiperda* (sf9; *see, e.g.,* Luckow et al. (1988) *Bio/technology 6:*47-55 and U.S. Patent No. 4,745,051).

TABLE 3

Plasmid(s) that Encode the Protein***	**Description of Fusion Protein**	**Fusion Protein Name**
N/A**	wild type FGF chemical conjugate*	CCFS1
N/A	mutein FGF C78S chemical conjugate*	CCFS2
N/A	mutein FGF C96S chemical conjugate*	CCFS3
N/A	mutein FGF C96S Cys-SAP CYS-1 chemical conjugate*	CCFS4
PZ1A, PZIB, PZIC, PZID, PZIE	wild type (FGF-Ala-Met-SAP) fusion protein**	FPFS1
PZ50B1	SAP CYS-1 PET 11a BL21(DE3)	FPS1
PZ51B1	SAP CYS+4 PET 11a BL21(DE3)	FPS2

PZ51E1	SAP CYS+4 PET 15b BL21(DE3)	FPS2
PZ52B1	SAP CYS+10 PET 11a BL21(DE3)	FPS3
PZ52E1	SAP CYS+10 PET 15b BL21(DE3)	FPS3
PZ30B1	HBEGF PET 11a BL21(DE3)	FPH1
PZ31B1	HBEGF-Val-Met-SAPPET11a BL21(DE3)	FPHS1
PZ32B1	HBEGF-Ala-Met-SAP PET11a BL21(DE3)	FPHS2
PZ33B1	HBEGF-Ala-Met--TRYPSIN-Ala-Met-SAP PET11a BL21(DE3)	FPHS3
PZ34B1	HBEGF-Ala-Met-CAT D-Ala-Met--SAP PET11a BL21(DE3)	FPHS4
PZ35B1	HBEGF-(Gly_4Ser)(Gly_2Ser)(Gly_4Ser)$_2$-SAP PET11aBL21(DE3)	FPHS5
PZ36B1	SAP-Ala-Met-Ala-HBEGF PET11a BL21(DE3)	FPSH1
PZ37B1	SAP-Ala-Met-(Gly_4Ser)$_4$-Ala-Met-Ala-HBEGF PET11a BL21(DE3)	FPSH2
PZ38I	Met-Cys HBEGF Viral Stock	FPH2
PZ39I	Met-Cys-Ala-Met-Ala-HBEGF Viral Stock	FPH3
PZ40I	Met-Cys-Ala-Met-(Gly_4Ser)$_2$-Ala-Met-Ala-HBEGF Viral Stock	FPH4
PZ41I	Met-Cys-Ala-Met-(Gly_4Ser)$_4$-Ala-Met-Ala-HBEGF Viral Stock	FPH5

* Details regarding these constructs are described in U.S. Application Serial Nos. 08/213,446 and 08/213,447, and PCT Appln. US 94/08511, filed July 27, 1994.

** N/A = not applicable

***The plasmids, such as PZ1A1 are designated with (i) a PZnumber (PZ1), followed by (ii) a letter (A), and optionally (iii) followed by a number (1). The key to these designations: (i) PZnumber - refers to the construct number, (ii) the next letter refers to the plasmid into which the construct was cloned, A=pET 11 wthout the T7 transcription terminator, B=pET 11 with the T7 transcription terminator, c=pET 13, D=pET 12, E=pET 15, F=λpPL, G=pKK 223-3, H=PRZ 1 (pKK223-3+Kan^R), I=pBlueBac III,

J=PRZ2 (pKK223-3 + Kan^R + lacI gene) and (ii) the optional number (or letter) refers to the bacterial strain (number) or insect host (letter) in which the plasmid was introduced, 1=BL21(DE3), 2=BL21(DE3)+pLYS S; 3=HMS174(DE3), 4=HMS174(DE3)+pLYS S, 5=N4830(cI8576) and 7=NovaBlue.

Fusion proteins FPHS5 and FPSH2 are purified from cell paste. Briefly, cell paste is suspended in 3-4 volumes of cell lysis buffer containing 10 mM sodium citrate, pH 6.0, 1 M urea, 5 mM EDTA, 5 mM EGTA and 50 mM NaCl. The lysate is passaged 3 times through a microfluidizer and diluted to 10 volumes with lysis buffer. The concentration of urea should be less than 8 M to reduce viscosity, and it is not necessary to include a cocktail of protease inhibitors. Urea is necessary for isolation of active protein. The extract is loaded onto an expanded bed of Streamline SP cation-exchange resin equilibrated with lysis buffer. Proteins are eluted with 2 buffers containing increasing NaCl concentrations: the first buffer contains 0.25 M NaCl and the second buffer contains 0.8 M NaCl. The second eluate is diluted in buffer without NaCl and subjected to anion-exchange chromatography on Q-Sepharose to remove DNA, endotoxins and contaminating proteins, and cation-exchange chromatography on SP-Sepharose to remove other contaminants. Proteins bound to the S-Sepharose column are eluted with a gradient of 0.25 to 1 M NaCl in buffer. Ammonium sulfate is added to the fusion proteins. As a positive selection, the protein is loaded onto a phenyl-Sepharose HP column and eluted with buffer containing 2 M ammonium sulfate. Monothioglycerol is added to the fusion protein. The protein is dialyzed and subjected to size -exclusion chromatography on S-100 resin. No heparin affinity chromatography is performed and a refolding protocol is not necessary to attain active material in the case of conjugates. It will be readily recognized that other equivalent resins and buffers may be readily substituted at each step in accordance with the purpose of each purification step. That is, for example, other equivalent cation exchange resins may be used in place of SP-Sepharose.

FPHS2 and FPHS1 fusion proteins are purified as above except that a heparin sepharose FF affinity column was additionally used prior to the S-100 column..

F. Properties and use of the chemical conjugates and fusion proteins

The conjugates provided herein can be used *in vitro* to identify cells, particularly tumor cells that express receptors to which the conjugate selectively binds and which internalizes the conjugates. The cells are contacted with the conjugates and assayed for proliferation. Cells in which proliferation is inhibited express receptors to which HBEGF binds. If such cells are derived from a tumor, such tumor will be a candidate for treatment with the HBEGF conjugate. If such cells are a cell line, such

cell line will be useful in drug screening assays for identification of compounds that modulate the activity of HBEGF receptors (*see, e.g.,* U.S. Patent Nos. 5,208,145, 5,071,773, 4,981,784, 4,603,106, which describe such assays for other receptors).

Each of the HBEGF-containing conjugates produced by the methods described herein can be tested, using a variety of well known *in vitro* and *in vivo* assays, for their ability to exert a cytopathic effect. For example the Promega (Madison, WI) CellTiter 96 Cell Proliferation/Cytotoxicity Assay described above and in Example 4 may be employed. In addition, *in vitro* cytotoxicity assays described in, for example, Kreitman et al. (1991) *Bioconjugate Chem. 3:*63-68; Epstein et al. (1991) *Circulation 84:*778-787, and the like, may be employed to test the conjugates produced herein.

In another assay that may be employed EGF-receptor expressing cells are plated in 96-well tissue culture plates at 1000-3000 cells/well in their respective medium. One day later, the medium is removed, and medium containing 1 pM to 1μM of the conjugate HBEGF-SAP, free SAP, free HBEGF, and HBEGF+SAP are added. Cells are treated in triplicate and maintained at 37^{O} C and 5% CO2. Forty-eight hours after the treatment is initiated, the MTT colorometric assay is utilized to measure cell sensitivity to HBEGF-SAP conjugates (Mossman, T. (1983) *J. Immunological Meth. 65:*55-63). Results are expressed as the mean optical density from treated wells, normalized to media controls, as a function of the HBEGF-SAP, free SAP, free HBEGF, and HBEGF+SAP concentration. The 50% inhibition values are calculated from dose-response curves and represent the concentration which resulted in a 50% reduction in cell number.

G. Formulation and administration of pharmaceutical compositions

As used herein, treatment means any manner in which the symptoms of a condition, disorder or disease are ameliorated or otherwise beneficially altered, reduced or relieved. Treatment also encompasses any pharmaceutical use of the compositions herein.

As used herein, amelioration of the symptoms of a particular disorder by administration of a particular pharmaceutical composition refers to any lessening, whether permanent or temporary, lasting or transient that can be attributed to or associated with administration of the composition.

The conjugates herein may be formulated into pharmaceutical compositions suitable for topical, local, intravenous and systemic application. Effective concentrations of one or more of the conjugates are mixed with a suitable pharmaceutical carrier or vehicle. The concentrations or amounts of the conjugates that are effective requires delivery of an amount, upon administration, that ameliorates the

symptoms or treats the disease. Typically, the compositions are formulated for single dosage administration. Therapeutically effective concentrations and amounts may be determined empirically by testing the conjugates in known *in vitro* and *in vivo* systems, such as those described here; dosages for humans or other animals may then be extrapolated therefrom.

Upon mixing or addition of the conjugate(s) with the vehicle, the resulting mixture may be a solution, suspension, emulsion or the like. The form of the resulting mixture depends upon a number of factors, including the intended mode of administration and the solubility of the conjugate in the selected carrier or vehicle. The effective concentration is sufficient for ameliorating the symptoms of the disease, disorder or condition treated and may be empirically determined based upon *in vitro* and/or *in vivo* data, such as the data from the mouse xenograft model for tumors or rabbit ophthalmic model. If necessary, pharmaceutically acceptable salts or other derivatives of the conjugates may be prepared.

Pharmaceutical carriers or vehicles suitable for administration of the conjugates provided herein include any such carriers known to those skilled in the art to be suitable for the particular mode of administration. In addition, the conjugates may be formulated as the sole pharmaceutically active ingredient in the composition or may be combined with other active ingredients.

As used herein, pharmaceutically acceptable salts, esters or other derivatives of the conjugates include any salts, esters or derivatives that may be readily prepared by those of skill in this art using known methods for such derivatization and that produce compounds that may be administered to animals or humans without substantial toxic effects and that either are pharmaceutically active or are prodrugs. A prodrug is a compound that, upon *in vivo* administration, is metabolized or otherwise converted to the biologically, pharmaceutically or therapeutically active form of the compound. To produce a prodrug, the pharmaceutically active compound is modified such that the active compound will be regenerated by metabolic processes. The prodrug may be designed to alter the metabolic stability or the transport characteristics of a drug, to mask side effects or toxicity, to improve the flavor of a drug or to alter other characteristics or properties of a drug. By virtue of knowledge of pharmacodynamic processes and drug metabolism *in vivo*, those of skill in this art, once a pharmaceutically active compound is known, can design prodrugs of the compound (*see, e.g.*, Nogrady (1985) *Medicinal Chemistry A Biochemical Approach*, Oxford University Press, New York, pages 388-392).

The conjugates can be administered by any appropriate route, for example, orally, parenterally, intravenously, intradermally, subcutaneously, or topically, in liquid, semi-liquid or solid form and are formulated in a manner suitable for each route of administration. Preferred modes of administration depend upon the indication treated. Dermatological and ophthalmologic indications will typically be treated locally; whereas, tumors and vascular proliferative disorders, will typically be treated by systemic, intradermal or intramuscular, modes of administration.

The conjugate is included in the pharmaceutically acceptable carrier in an amount sufficient to exert a therapeutically useful effect in the absence of undesirable side effects on the patient treated. It is understood that number and degree of side effects depends upon the condition for which the conjugates are administered. For example, certain toxic and undesirable side effects are tolerated when treating life-threatening illnesses, such as tumors, that would not be tolerated when treating disorders of lesser consequence.

The concentration of conjugate in the composition will depend on absorption, inactivation and excretion rates thereof, the dosage schedule, and amount administered as well as other factors known to those of skill in the art.

As used herein an effective amount of a compound for treating a particular disease is an amount that is sufficient to ameliorate, or in some manner reduce the symptoms associated with the disease. Such amount may be administered as a single dosage or may be administered according to a regimen, whereby it is effective. The amount may cure the disease but, typically, is administered in order to ameliorate the symptoms of the disease. Repeated administration may be required to achieve the desired amelioration of symptoms.

Typically a therapeutically effective dosage should produce a serum concentration of active ingredient of from about 0.1 ng/ml to about 50-100 μg/ml. The pharmaceutical compositions typically should provide a dosage of from about 0.01 mg to about 100 - 2000 mg of conjugate, depending upon the conjugate selected, per kilogram of body weight per day. Typically, for intravenous or systemic treatment a daily dosage of about between 0.05 and 0.5 mg/kg should be sufficient and can be administered as a bolus or continuous infusion. Local application for ophthalmic disorders should provide about 1 ng up to 100 μg, per single dosage administration. It is understood that the amount to administer will be a function of the conjugate selected, the indication treated, and possibly the side effects that will be tolerated. Dosages can be empirically determined using recognized models for each disorder.

The active ingredient may be administered at once, or may be divided into a number of smaller doses to be administered at intervals of time. It is understood that the precise dosage and duration of treatment is a function of the disease being treated and may be determined empirically using known testing protocols or by extrapolation from *in vivo* or *in vitro* test data. It is to be noted that concentrations and dosage values may also vary with the severity of the condition to be alleviated. It is to be further understood that for any particular subject, specific dosage regimens should be adjusted over time according to the individual need and the professional judgment of the person administering or supervising the administration of the compositions, and that the concentration ranges set forth herein are exemplary only and are not intended to limit the scope or practice of the claimed compositions.

Solutions or suspensions used for parenteral, intradermal, subcutaneous, or topical application can include any of the following components: a sterile diluent, such as water for injection, saline solution, fixed oil, polyethylene glycol, glycerine, propylene glycol or other synthetic solvent; antimicrobial agents, such as benzyl alcohol and methyl parabens; antioxidants, such as ascorbic acid and sodium bisulfite; chelating agents, such as ethylenediaminetetraacetic acid (EDTA); buffers, such as acetates, citrates and phosphates; and agents for the adjustment of tonicity such as sodium chloride or dextrose. Parental preparations can be enclosed in ampules, disposable syringes or multiple dose vials made of glass, plastic or other suitable material.

If administered intravenously, suitable carriers include physiological saline or phosphate buffered saline (PBS), and solutions containing thickening and solubilizing agents, such as glucose, polyethylene glycol, and polypropylene glycol and mixtures thereof. Liposomal suspensions may also be suitable as pharmaceutically acceptable carriers. These may be prepared according to methods known to those skilled in the art.

The conjugates may be prepared with carriers that protect them against rapid elimination from the body, such as time release formulations or coatings. Such carriers include controlled release formulations, such as, but not limited to, implants and microencapsulated delivery systems, and biodegradable, biocompatible polymers, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, polyorthoesters, polylacetic acid and others. These are particularly useful for application to the eye for ophthalmic indications following or during surgery in which only a single administration is possible. Methods for preparation of such formulations are known to those skilled in the art.

The conjugates may be formulated for local or topical application, such as for topical application to the skin and mucous membranes, such as in the eye, in the form of gels, creams, and lotions and for application to the eye or for intracisternal or intraspinal application. Such solutions, particularly those intended for ophthalmic use, may be formulated as 0.01% -10% isotonic solutions, pH about 5-7, with appropriate salts. The ophthalmic compositions may also include additional components, such as hyaluronic acid. The conjugates may be formulated as aerosols for topical application (*see, e.g.*, U.S. Patent Nos. 4,044,126, 4,414,209, and 4,364,923).

If oral administration is desired, the conjugate should be provided in a composition that protects it from the acidic environment of the stomach. For example, the composition can be formulated in an enteric coating that maintains its integrity in the stomach and releases the active compound in the intestine. The composition may also be formulated in combination with an antacid or other such ingredient.

Oral compositions will generally include an inert diluent or an edible carrier and may be compressed into tablets or enclosed in gelatin capsules. For the purpose of oral therapeutic administration, the active compound or compounds can be incorporated with excipients and used in the form of tablets, capsules or troches. Pharmaceutically compatible binding agents and adjuvant materials can be included as part of the composition.

The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder, such as microcrystalline cellulose, gum tragacanth and gelatin; an excipient such as starch and lactose, a disintegrating agent such as, but not limited to, alginic acid and corn starch; a lubricant such as, but not limited to, magnesium stearate; a glidant, such as, but not limited to, colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; and a flavoring agent such as peppermint, methyl salicylate, and fruit flavoring.

When the dosage unit form is a capsule, it can contain, in addition to material of the above type, a liquid carrier such as a fatty oil. In addition, dosage unit forms can contain various other materials which modify the physical form of the dosage unit, for example, coatings of sugar and other enteric agents. The conjugates can also be administered as a component of an elixir, suspension, syrup, wafer, chewing gum or the like. A syrup may contain, in addition to the active compounds, sucrose as a sweetening agent and certain preservatives, dyes and colorings and flavors.

The active materials can also be mixed with other active materials that do not impair the desired action, or with materials that supplement the desired action, such as cis-platin for treatment of tumors.

Finally, the compounds may be packaged as articles of manufacture containing packaging material, one or more conjugates or compositions as provided herein within the packaging material, and a label that indicates the indication for which the conjugate is provided.

H. Therapeutic uses of the HBEGF conjugates

The conjugates provided herein can be used in pharmaceutical compositions to treat HBEGF-mediated pathophysiological conditions by targeting to cells that bear HBEGF receptors and inhibiting proliferation of or causing death of the cells. Such pathophysiological conditions include, for example, certain tumors, such as renal cell carcinomas and breast and bladder tumors, psoriasis, ophthalmic disorders involving epithelial cells, such as recurrence of pterygii and secondary lens clouding. The treatment is effected by administering a therapeutically effective amount of the HBEGF conjugate, for example, in a physiological vehicle suitable for local or systemic application. In particular, for treatment of localized skin disorders the conjugate is formulated for topical, local or intralesional application to the skin and is applied topically, locally or intralesional.

1. Treatment of pathophysiological smooth muscle cell proliferation

Atherosclerosis, which results from the development of an intimal lesion and the subsequent narrowing of the vessel lumen, commonly results from the buildup of plaque which lines the interior of blood vessels, particularly the arteries. In recent years, a number of surgical procedures have been developed that interarterially remove such plaque, often by balloon catheterization or other such treatments, either by compressing it against or scraping it away from the interior surface of the artery. Not infrequently, the patient so treated experiences a recurrence of narrowing of the vessel lumen in a relatively short period thereafter. This narrowing following treatment to remove plaque is referred to as restenosis.

Methods are provided herein for treating restenosis by administering an effective amount of an HBEGF cytotoxic conjugate, such that the HBEGF conjugate inhibits smooth muscle cell proliferation in the lining of vessels that have been injured without inhibiting proliferation of endothelial cells that is necessary for preventing or treating restenosis following vascular injury. It can be administered locally or intravenously. A medicament containing an HBEGF-toxin, preferably saporin, conjugate will be targeted to proliferating smooth muscle cells in the treated arteries and relatively few infusions (or a few, *i.e.*, up to about 3-5) should prevent restenosis.

Preferably, the medicament containing the conjugate is administered intravenously (IV), although treatment by localized administration of the conjugate may

be tolerated in some instances. Generally, the medicament containing the conjugate is injected into the circulatory system of a patient in order to deliver a dose of cytotoxin to the targeted cells by first binding the conjugate to high HBEGF receptors expressed by such cells.

The efficiency with which a cytotoxin, such as saporin or a Ricin A chain or a similar RIP, can inhibit protein synthesis and consequently interfere with DNA synthesis is fairly widely known. Accordingly, the dosage of the conjugate that is administered will, to some extent, depend upon the particular cytotoxin chosen; however, doses of the conjugate in the general range of about 0.01 mg to about 100 mg of the conjugate per kilogram of body weight are expected to be employed as a daily dosage. There may be particular advantages in administering a daily dosage of about 0.1 mg/kg (i.e. between 0.05 and 0.3 mg/kg).

2. Treatment of tumors

Tumors, particularly solid tumors, including bladder, breast, ovarian, pancreatic and some colon carcinomas, have receptors to which HBEGFs bind. The susceptibility of particular tumors can be ascertained by isolating the cell, and contacting them with an HBEGF cytotoxic conjugate and determining sensitivity to the conjugate by a standard proliferation assay. This should identify those tumors that would be amenable to treatment and also identifies tumor cells that express receptors to which HBEGF binds. Cytotoxic conjugates, such as HBEGF conjugated with the saporin molecule (HBEGF-SAP), are inhibitors of cell growth *in vitro* for cell lines that express HBEGF receptors. Such *in vitro* activity should be extrapolatable to *in vivo* activity. In *vivo* activity may be assessed using recognized animal models, such as the mouse xenograft model for anti-tumor activity (*see, e.g.*, Beitz et al. (1992) *Cancer Research 52*:227-230; Houghton et al. (1982) *Cancer Res. 42*:535-539; Bogden et al. (1981) *Cancer (Philadelphia) 48*:10-20; Hoogenhout et al. (1983) *Int. J. Radiat. Oncol., Biol. Phys. 9*:871-879; Stastny et al. (1993) *Cancer Res. 53*:5740-5744). Cell lines that are sensitive to the cytotoxic HBEGF conjugates can be grown subcutaneously as solid tumor xenografts in nude mice, and administration of HBEGF-SAP conjugates to such mice should show rapid reduction in tumor volume in those cell lines which responded to treatment of the conjugate.

Treatment of mammals, including human patients, would be similarly effected by administering a therapeutically effective amount of the HBEGF conjugate in a physiologically acceptable carrier. Specifically, in the treatment, the conjugates are used to target cytotoxic agents to human solid tumors, including bladder or breast

tumors, to inhibit the proliferation of such cells. The conjugates are also used to target HBEGF receptor-expressing cells in similar tumorigenic pathophysiological conditions.

The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

EXAMPLE 1

RECOMBINANT PRODUCTION OF SAPORIN

The manipulations described in this example are also described in International PCT Application WO 93/25688 and copending U.S. Applications Serial Nos. 08/145,829 and 07/918,718.

A. Materials and methods

1. Reagents

Restriction and modification enzymes were purchased from BRL (Gaithersburg, MD), Stratagene (La Jolla, CA) and New England Biolabs (Beverly, MA). Native SAP was obtained from *Saponaria officinalis* (*see, e.g.*, Stirpe et al. (1983) *Biochem. J. 216:*617-625). Briefly, the seeds were extracted by grinding in 5 mM sodium phosphate buffer, pH 7.2 containing 0.14 M NaCl, straining the extracts through cheesecloth, followed by centrifugation at 28,000 g for 30 min to produce a crude extract, which was dialyzed against 5 mM sodium phosphate buffer, pH 6.5, centrifuged and applied to CM-cellulose (CM 52, Whatman, Maidstone, Kent, U.K.). The CM column was washed and SO-6 was eluted with a 0-0.3 M NaCl gradient in the phosphate buffer.

2. Bacterial Strains

E. coli strain JA221 (lpp$^-$ hdsM+ trpE5 leuB6 lacY recA1 F'[lacI^q lac$^+$ pro$^+$]) is publicly available from the American Type Culture Collection (ATCC), Rockville, MD 20852, under the accession number ATCC 33875. (JA221 is also available from the Northern Regional Research Center (NRRL), Agricultural Research Service, U.S. Department of Agriculture, Peoria, IL 61604, under the accession number NRRL B-15211; see, also, U.S. Patent No. 4,757,013 to Inouye; and Nakamura et al. (1979) *Cell 18:*1109-1117). Strain INV1α is commercially available from Invitrogen, San Diego, CA; and strains Novablue and BL21(DE3) are commercially available (Novagen, Madison WI).

3. DNA Manipulations

The restriction and modification enzymes employed herein are commercially available in the U.S. Native saporin and rabbit polyclonal antiserum to saporin were obtained as previously described in Lappi et al. (1985) *Biochem. Biophys. Res. Comm. 129:*934-942. Ricin A chain is commercially available from SIGMA, Milwaukee, WI. Antiserum was linked to Affi-gel 10 (BIO-RAD, Emeryville, CA) according to the manufacturer's instructions. Sequencing was performed using the

Sequenase kit of United States Biochemical Corporation (version 2.0) according to the manufacturer's instructions. Minipreparation and maxipreparation of plasmids, preparation of competent cells, transformation, M13 manipulation, bacterial media, Western blotting, and ELISA assays were according to Sambrook et al. ((1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). Purification of DNA fragments was achieved using the Geneclean II kit, purchased from Bio 101 (La Jolla, CA).

4. Sodium dodecyl sulfate (SDS) gel electrophoresis and Western blotting.

SDS gel electrophoresis was performed on a PhastSystem utilizing 20% gels (Pharmacia). Western blotting was accomplished by transfer of electrophoresed protein to nitrocellulose using the PhastTransfer system (Pharmacia), as described by the manufacturer. The antiserum to SAP was used at a dilution of 1:1000. Horseradish peroxidase labeled anti-IgG was used as the second antibody as described (Davis, L., Dibner et al. (1986) *Basic Methods in Molecular Biology*, pp. 1-338, Elsevier Science Publishing Co., New York).

5. Cell-free assay for cytotoxic activity

The RIP activity of saporin can be and is determined in an *in vitro* assay measuring cell-free protein synthesis in a nuclease-treated rabbit reticulocyte lysate (Promega). Samples of saporin are added on ice to 35 µl of rabbit reticulocyte lysate and 10 µl of a reaction mixture containing 0.5 µl of Brome Mosaic Virus RNA, 1 mM amino acid mixture minus leucine, 5 µCi of tritiated leucine and 3 µl of water. Assay tubes are incubated 1 hour in a 30 C water bath. The reaction is stopped by transferring the tubes to ice and adding 5 µl of the assay mixture, in triplicate, to 75 µl of 1 N sodium hydroxide, 2.5% hydrogen peroxide in the wells of a Millititer HA 96-well filtration plate (Millipore). When the red color has bleached from the samples, 300 µl of ice cold 25% trichloroacetic acid (TCA) are added to each well and the plate left on ice for another 30 min. Vacuum filtration is performed with a Millipore vacuum holder. The wells are washed three times with 300 µl of ice cold 8% TCA. After drying, the filter paper circles are punched out of the 96-well plate and counted by liquid scintillation techniques. The IC_{50} for recombinant and native saporin is approximately 20 pM.

B. Isolation of DNA encoding saporin

1. Isolation of genomic DNA and preparation of amplification primers

Saponaria officinalis leaf genomic DNA was prepared as described in Bianchi et al. (1988) *Plant Mol. Biol. 11*:203-214. Primers for genomic DNA

amplifications were synthesized in a 380B automatic DNA synthesizer. The primer corresponding to the "sense" strand of saporin (SEQ ID NO. 27 includes an *Eco*R I restriction site adapter immediately upstream of the DNA codon for amino acid -15 of the native saporin N-terminal leader sequence (SEQ ID NO. 27): 5'-CTGCAGAATTCGCATGGATCCTGCTTCAAT-3'. The primer 5'-CTGCAGAATTCGCCTCGTTTGACTACTTTG-3' (SEQ ID NO. 28) corresponds to the "antisense" strand of saporin and complements the coding sequence of saporin starting from the last 5 nucleotides of the DNA encoding the carboxyl end of the mature peptide. Use of this primer introduced a translation stop codon and an *Eco*RI restriction site after the sequence encoding mature saporin.

2. Amplification of DNA encoding saporin

Unfractionated *Saponaria officinalis* leaf genomic DNA (1 µl) was mixed in a final volume of 100 µl containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 0.01% gelatin, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.8 µg of each primer. Next, 2.5 U TaqI DNA polymerase (Perkin Elmer Cetus) was added and the mixture was overlaid with 30 µl of mineral oil (Sigma). Incubations were done in a DNA Thermal Cycler (Ericomp). One cycle included a denaturation step (94 C for 1 min.), an annealing step (60 C for 2 min.), and an elongation step (72 C for 3 min.). After 30 cycles, a 10 µl aliquot of each reaction was run on a 1.5% agarose gel to verify the correct size of the amplified product.

The amplified DNA was digested with *Eco*RI and subcloned into *Eco*RI I-restricted M13mp18 (NEW ENGLAND BIOLABS, Beverly, MA; see, also, Yanisch-Perron et al. (1985), "Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors", *Gene 33:*103). Single-stranded DNA from recombinant phages was sequenced using oligonucleotides based on internal points in the coding sequence of saporin (see, Bennati et al. (1989) *Eur. J. Biochem. 183:*465-470). Nine of the M13mp18 derivatives were sequenced and compared. Of the nine sequenced clones, five had unique sequences, set forth as SEQ ID NOs. 8-12, respectively. The clones were designated M13mp18-G4, -G1, -G2, -G7, and -G9. Each of these clones contains all of the saporin coding sequence and 45 nucleotides of DNA encoding the native saporin N-terminal leader peptide.

C. pOMPAG4 Plasmid Construction

M13 mp18-G4, containing the clone containing saporin of SEQ ID NO. 8 from Example 1.B., was digested with *Eco*R I, and the resulting fragment was ligated into the *Eco*R I site of the vector pIN-IIIompA2 (*see, e.g.*, see, U.S. Patent NO 4,575,013 to Inouye; and Duffaud et al. (1987) *Meth. Enz. 153:*492-507) using the

methods described in Example 1.A. The ligation was accomplished such that the DNA encoding saporin, including the N-terminal extension, was fused to the leader peptide segment of the bacterial ompA gene. The resulting plasmid pOMPAG4 contains the lpp promoter (Nakamura et al. (1979) *Cell 18:*1109-1117), the *E. coli* lac promoter operator sequence (lac O) and the *E. coli* ompA gene secretion signal in operative association with each other and with the saporin and native N-terminal leader-encoding DNA listed in SEQ ID NO 8. The plasmid also includes the *E. coli* lac repressor gene (lac I).

The M13 mp18-G1, -G2, -G7, and -G9 clones obtained from Example 1.B.2, containing SEQ ID NOs. 9-12, respectively, are digested with *Eco*R I and ligated into *Eco*R I digested pIN-IIIompA2 as described for M13 mp18-G4 above in this example. The resulting plasmids, labeled pOMPAG1, pOMPAG2, pOMPAG7, pOMPA9, are screened, expressed, purified, and characterized as described for the plasmid pOMPAG4.

INV1α competent cells were transformed with pOMPAG4 and cultures containing the desired plasmid structure were grown further in order to obtain a large preparation of isolated pOMPAG4 plasmid using methods described in Example 1.A.

D. Saporin expression in *E. coli*

The pOMPAG4 transformed *E. coli* cells were grown under conditions in which the expression of the saporin-containing protein is repressed by the lac repressor to an O.D. in or at the end of the log phase of growth after which IPTG was added to induce expression of the saporin-encoding DNA.

To generate a large-batch culture of pOMPAG4 transformed *E. coli* cells, an overnight culture (lasting approximately 16 hours) of JA221 *E. coli* cells transformed with the plasmid pOMPAG4 in LB broth (*see, e.g.*, Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY) containing 125 mg/ml ampicillin was diluted 1:100 into a flask containing 750 ml LB broth with 125 mg/ml ampicillin. Cells were grown to logarithmic phase with shaking at 37^{0} C until the optical density at 550 nm reached 0.9.

In the second step, saporin expression was induced by the addition of IPTG (Sigma) to a final concentration of 0.2 mM. Induced cultures were grown for 2 additional hours and then harvested by centrifugation (25 min., 6500 x g). The cell pellet was resuspended in ice cold 1.0 M TRIS, pH 9.0, 2 mM EDTA (10 ml were added to each gram of pellet). The resuspended material was kept on ice for 20-60 minutes and then centrifuged (20 min., 6500 x g) to separate the periplasmic fraction of

E. coli, which corresponds to the supernatant, from the intracellular fraction corresponding to the pellet.

As described below, (see, Example 3), it has been found that it is preferable to perform the manipulations previously conducted at 37° C at 30° C.

E. Assay for cytotoxic activity

The RIP activity of recombinant saporin was compared to the RIP activity of native SAP in an *in vitro* assay measuring cell-free protein synthesis in a nuclease-treated rabbit reticulocyte lysate (Promega). Samples of immunoaffinity-purified saporin (*see, e.g.*, Lappi et al. (1985) *Biochem. Biophys. Res. Comm. 129*:934-942) were diluted in PBS and 5 µl of sample was added on ice to 35 µl of rabbit reticulocyte lysate and 10 µl of a reaction mixture containing 0.5 µl of Brome Mosaic Virus RNA, 1 mM amino acid mixture minus leucine, 5 µCi of tritiated leucine and 3 µl of water. Assay tubes were incubated 1 hour in a 30 C water bath. The reaction was stopped by transferring the tubes to ice and adding 5 µl of the assay mixture, in triplicate, to 75 µl of 1 N sodium hydroxide, 2.5% hydrogen peroxide in the wells of a Millititer HA 96-well filtration plate (Millipore). When the red color had bleached from the samples, 300 µl of ice cold 25% trichloroacetic acid (TCA) were added to each well and the plate left on ice for another 30 min. Vacuum filtration was performed with a Millipore vacuum holder. The wells were washed three times with 300 µl of ice cold 8% TCA. After drying, the filter paper circles were punched out of the 96-well plate and counted by liquid scintillation techniques.

The IC_{50} for the recombinant and native saporin were approximately 20 pM. Therefore, recombinant saporin-containing protein has full protein synthesis inhibition activity when compared to native saporin.

EXAMPLE 2

PREPARATION OF STARTING PLASMIDS - PZIA, PZIB, PZIC AND PZID

A. General Descriptions

1. Bacterial Strains and Plasmids:

E. coli strains BL21(DE3), BL21(DE3)pLysS, HMS174(DE3) and HMS174(DE3)pLysS were purchased from Novagen, Madison, WI. Plasmid pFC80, described below, has been described in the WIPO International Patent Application No. WO 90/02800, except that the bFGF coding sequence in the plasmid designated pFC80

herein has the sequence set forth as SEQ ID NO. 34, nucleotides 1-465. The plasmids described herein may be prepared using pFC80 as a starting material or, alternatively, by starting with a fragment containing the CII ribosome binding site (SEQ ID NO. 19) linked to the FGF-encoding DNA (SEQ ID NO. 34).

2. DNA Manipulations

The restriction and modification enzymes employed here are commercially available in the U.S. Native SAP, chemically conjugated bFGF-SAP and rabbit polyclonal antiserum to SAP were obtained as described, for example, in Lappi et al. (1985) *Biochem. Biophys. Res. Comm. 129*:934-942, Lappi et al. (1989) *Biochem. Biophys., Res. Comm. 160*:917-923 and U.S. Patent No. 5,191,067. The pET System Induction Control was purchased from Novagen, Madison, WI. The sequencing of the different constructions was done using the Sequenase kit of United States Biochemical Corporation (version 2.0). Minipreparation and maxipreparations of plasmids, preparation of competent cells, transformation, M13 manipulation, bacterial media and Western blotting were performed using routine methods (*see, e.g.,*.Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). The purification of DNA fragments was done using the Geneclean II kit, purchased from Bio 101. SDS gel electrophoresis was performed on a Phastsystem (Pharmacia).

3. Cytotoxicity assays of conjugates.

Cytotoxicity experiments are performed with the Promega (Madison, WI) CellTiter 96 Cell Proliferation/Cytotoxicity Assay. Cell types are A431 or SK-MEL28 cells. 2500 cells are plated per well.

B. Construction of plasmids encoding FGF-SAP fusion proteins

1. Construction of FGFM13 that contains DNA encoding the CI ribosome binding site linked to FGF

A *Nco* I restriction site was introduced into the SAP-encoding DNA of the M13mp18-G4 clone, described in Example 1, by site-directed mutagenesis method using the Amersham *In vitro*-mutagenesis system 2.1. The oligonucleotide employed to create the *Nco* I restriction site was synthesized using a 380B automatic DNA synthesizer (Applied Biosystems) and is has the sequence (SEQ ID NO. 17): CAACAACTGCCATGGTCACATC. This oligonucleotide containing the Nco I site replaced the original SAP-containing coding sequence at SEQ ID NO. 8, nts 32-53. The resulting M13mp18-G4 derivative was designated mpNG4.

In order to produce a bFGF coding sequence in which the stop codon was removed, the FGF-encoding DNA was subcloned into a M13 phage and subjected

to site-directed mutagenesis. Plasmid pFC80 is a derivative of pDS20 (*see, e.g.*, Duester et al. (1982) *Cell 30:*855-864; *see also* U.S. Patent Nos. 4,914,027, 5,037,744, 5,100,784, and 5,187,261; see, also, PCT International Application No. WO 90/02800; and European Patent Application No. EP 267703 A1), which is almost the same as plasmid pKG1800 (see, Bernardi et al. (1990) *DNA Sequence 1:*147-150; see, also McKenney et al. (1981) pp. 383-415 in *Gene Amplification and Analysis 2: Analysis of Nucleic Acids by Enzymatic Methods* Chirikjian et al., eds, North Holland Publishing Company, Amsterdam) except that it contains an extra 440 bp at the distal end of *galK* between nucleotides 2440 and 2880 in pDS20. Plasmid pKG1800 includes the 2880 bp EcoR I-Pvu II of pBR322 that contains the ampicillin resistance gene and an origin of replication.

Plasmid pFC80 was prepared from pDS20 by replacing the entire *galK* gene with the FGF-encoding DNA of SEQ ID NO. 34, inserting the trp promoter (SEQ ID NO. 18) and the bacteriophage lambda CII ribosome binding site (SEQ. ID NO. 19; *see, e.g.*, Schwarz et al. (1978) *Nature 272:*410) upstream of and operatively linked to the FGF-encoding DNA. The Trp promoter can be obtained from plasmid pDR720 (Pharmacia PL Biochemicals) or synthesized according to SEQ ID NO. 18. Plasmid pFC80, contains the 2880 bp *Eco*R I-*Bam*H I fragment of plasmid pSD20, a synthetic *Sal* I-*Nde* I fragment that encodes the Trp promoter region (SEQ ID NO. 18):

*Eco*R I
AATTCCCCTGTTGACAATTAATCATCGAACTAGTTAACTAGTACGCAGCTTGGCTGCAG
and the CII ribosome binding site (SEQ ID NO.19):

Sal I Nde I
GTCGACCAAGCTTGGGCATACATTCAATCAATTGTTATCTAAGGAAATACTTACATATG.

The FGF-encoding DNA was removed from pFC80 by treating it as follows. The pFC80 plasmid was digested by Hga I and Sal I, which produces a fragment containing the CII ribosome binding site linked to the FGF-encoding DNA. The resulting fragment was blunt ended with DNA pol I (Klenow fragment) and inserted into M13mp18 that had been opened by Sma I and treated with alkaline phosphatase for blunt-end ligation. In order to remove the stop codon, an insert in the ORI minus direction was mutagenized using the Amersham kit, as described above, using the following oligonucleotide (SEQ ID NO. 20): GCTAAGAGCGCCATGGAGA, which contains 1 nucleotide between the FGF

carboxy terminal serine codon and a Nco I restriction site, and it replaced the following wild type FGF encoding DNA having SEQ ID NO. 21:

```
GCT AAG AGC TGA  CCA TGG AGA.
Ala Lys Ser STOP Pro Trp Arg
```

The resulting mutant derivative of M13mp18, lacking a native stop codon after the carboxy terminal serine codon of bFGF, was designated FGFM13. The mutagenized region of FGFM13 contained the correct sequence (SEQ ID NO. 22).

2. Preparation of plasmids pFS92 (PZ1A), PZ1B and PZ1C that encode the FGF-SAP fusion protein

a. Plasmid pFS92 (also designated PZ1A)

Plasmid FGFM13 was cut with *Nco* I and *Sac* I to yield a fragment containing the CII ribosome binding site linked to the bFGF coding sequence with the stop codon replaced.

The M13mp18 derivative mpNG4 containing the saporin coding sequence was also cut with restriction endonucleases *Nco* I and *Sac* I, and the bFGF coding fragment from FGFM13 was inserted by ligation to DNA encoding the fusion protein bFGF-SAP into the M13mp18 derivative to produce mpFGF-SAP, which contains the CII ribosome binding site linked to the FGF-SAP fusion gene. The sequence of the fusion gene is set forth in SEQ ID NO. 34 and indicates that the FGF protein carboxy terminus and the saporin protein amino terminus are separated by 6 nucleotides (SEQ ID NOs. 34 and 35, nts 466-471) that encode two amino acids Ala Met.

Plasmid FGF-SAP was digested with *Xba* I and *Eco*R I and the resulting fragment containing the bFGF-SAP coding sequence was isolated and ligated into plasmid pET 11a (available from Novagen, Madison, WI; for a description of the plasmids see U.S. Patent No. 4,952.496; see, also Studier et al. (1990) *Meth. Enz. 185:*60-89; Studier et al. (1986) *J. Mol. Biol. 189:*113-130; Rosenberg et al. (1987) *Gene 56:*125-135) that had also been treated with *Eco*R I and *Xba* I. The resulting plasmid was designated pFS92. It was renamed PZ1A.

Plasmid pFS92 (or PZ1A) contains DNA encoding the entire basic FGF protein (SEQ ID NO. 34), a 2-amino acid long connecting peptide, and amino acids 1 to 253 of the mature SAP protein. Plasmid pFS92 also includes the CII ribosome binding site linked to the FGF-SAP fusion protein and the T7 promoter region from pET 11a.

E. coli strain BL21(DE3)pLysS (Novagen, Madison WI) was transformed with pFS92 according to manufacturer's instructions and the methods described in Example 2.A.2.

b. Plasmid PZ1B

Plasmid pFS92 was digested with *Eco*R I, the ends repaired by adding nucleoside triphosphates and Klenow DNA polymerase, and then digested with *Nde* I to release the FGF-encoding DNA without the cII ribosome binding site. This fragment was ligated into pET 11a, which had been *Bam*H I digested, treated to repair the ends, and digested with *Nde* I. The resulting plasmid was designated PZ1B. PZ1B includes the T7 transcription terminator and the pET 11a ribosome binding site.

E. coli strain BL21(DE3) (Novagen, Madison WI) was transformed with PZ1B according to manufacturer's instructions and the methods described in Example 2.A.2.

c. Plasmid PZ1C

Plasmid PZ1C was prepared similarly to PZ1B but contains a kanamycin resistance gene and is based on the pET 13a vector.

d. Plasmid PZ1D

Plasmid pFS92 was digested with *Eco*R I and *Nde* I to release the FGF-encoding DNA without the CII ribosome binding site and the ends were repaired. This fragment was ligated into pET 12a, which had been *Bam*H I digested and treated to repair the ends. The resulting plasmid was designated PZ1D. PZ1D includes DNA encoding the OMP T secretion signal operatively linked to DNA encoding the fusion protein.

E. coli strains BL21(DE3), BL21(DE3)pLysS, HMS174(DE3) and HMS174(DE3)pLysS (Novagen, Madison WI) were transformed with PZ1D according to manufacturer's instructions and the methods described in Example 2.

EXAMPLE 3

PREPARATION OF MODIFIED SAPORIN

Saporin was modified by addition of a cysteine residue at the N-terminus-encoding portion of the DNA or by the addition of a cysteine at position 4 or 10. The resulting saporin is then reacted with an available cysteine or sulfhydryl-

reacting moiety on a targeting agent to produce conjugates that are linked via the added Cys or Met-Cys on saporin.

Modified SAP has been prepared by altering the DNA encoding the SAP by inserting DNA encoding Met-Cys at position -1 or by replacing the Ile or the Asn codon within 10 or fewer residues of the N-terminus with Cys. The resulting DNA has been inserted into pET 11a and pET 15b and expressed in BL21(DE3) cells. The resulting saporin proteins are designated FPS1 (saporin with Cys at -1), FPS2 (saporin with Cys at position 4) and FPS3 (saporin with Cys at position 10). A plasmid that encodes FPS1 and that has been used for expression of FPS1 has been designated PZ50B. Plasmids that encode FPS2 and that have been used for expression of FPS2 have been designated PZ51B (pET11a-based plasmid) and PZ51E (pET15b-based plasmid). Plasmids that encode FPS3 and that have been used for expression of FPS3 have been designated PZ52B (pET11a-based plasmid) and PZ52E (pET 15b-based plasmid).

A. Materials and Methods

1. Bacterial strains

Novablue (Novagen, Madison, WI) and BL21(DE3) (Novagen, Madison WI).

2. DNA manipulations

DNA manipulations were performed as described in Examples 1 and 2. Plasmid PZ1B (designated PZ1B1 (the "1" at the end refers to the bacterial host strain, BL21(DE3)) described in Example 2 was used as the DNA template.

B. Preparation of saporin with an added cysteine residue at the N-terminus

1. Primers

(a) Primer #1 corresponding to the sense strand of saporin, nucleotides 472-492 of SEQ ID NO. 34, incorporates a *NdeI* site and adds a cys codon 5' to the first codon of the mature protein (between Met and Val):

CATATGTGTGTCACATCAATCACATTAGAT (SEQ ID NO. 15).

(b) Primer #2 - Antisense primer complements the coding sequence of saporin spanning nucleotides 547-567 of SEQ ID NO. 34 and contains a *BamHI* site:

CAGGTTTGGATCCTTTACGTT (SEQ ID NO. 16).

2. Isolation of saporin-encoding DNA

PZIB DNA was amplified as follows using the above primers. PZ1B DNA (1 µl) was mixed in a final volume of 100 µl containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 0.01% gelatin, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.8 µg of each primer. Next, 2.5 U TaqI DNA polymerase (Boehringer Mannheim) was added and the mixture was overlaid with 30 µl of mineral oil (Sigma). Incubations were done in a DNA Thermal Cycler (Ericomp). One cycle included a denaturation step (94 C for 1 min), an annealing step (60 C for 2 min.), and an elongation step (72° C for 3 min). After 35 cycles, a 10 µl aliquot of each reaction was run on a 1.5% agarose gel to verify the correct size of the amplified product.

The amplified DNA was gel purified and digested with *Nde*I and *Bam*HI and subcloned into *Nde*I and *Bam*HI-digested pZ1B. This digestion and subcloning step removed the FGF-encoding DNA and 5' portion of SAP up to the *Bam*HI site at nucleotides 555-560 (SEQ ID NO. 34) and replaced this portion with DNA encoding a saporin molecule that contains a cysteine residue at position -1 relative to the start site of the native mature SAP protein (see, SEQ ID NO. 58). The resulting plasmid is designated pZ50B.

C. Preparation of saporin with a cysteine residue at position 4 or 10 of the native protein

These constructs were designed to introduce a cysteine residue at position 4 or 10 of the native protein by replacing the isoleucine residue at position 4 or the asparagine residue at position 10 with cysteine.

1. Materials

(a) Bacterial strains

The bacterial strains were Novablue and BL21(DE3) (Novagen, Madison, WI).

(b) DNA manipulations

DNA manipulations as described above.

2. Preparation of modified SAP-encoding DNA

SAP was amplified by polymerase chain reaction (PCR) from the parental plasmid pZ1B encoding the FGF-SAP fusion protein.

(a) Primers

(1) The primer corresponding to the sense strand of saporin, spanning nucleotides 466-501 of SEQ ID NO. 34, incorporates a *Nde*I site and replaces the Ile codon with a Cys codon at position 4 of the mature protein (SEQ ID NO. 59):

CATATGGTCACATCATGTACATTAGATCTAGTAAAT.

(2) The primer corresponding to the sense strand of saporin, nucleotides 466-515 of SEQ ID NO. 34, incorporates a *Nde*I site and replaces the Asn codon with a cys codon at position 10 of the mature protein (SEQ ID NO. 60)

CATATGGTCACATCAATCACATTAGATCTAGTATGTCCGACCGCGGGTCA.

(3) Primer #2 - Antisense primer complements the coding sequence of saporin spanning nucleotides 547-567 of SEQ ID NO. 34 and contains a BamHI site (SEQ ID NO. 16):

CAGGTTTGGATCCTTTACGTT.

(b) Amplification

The nucleic acid amplification reactions were performed as described above, using the following cycles: denaturation step 94°C for 1 min, annealing for 2 min at 60°C, and extension for 2 min at 72°C for 35 cycles. The amplified DNA was gel purified, digested with NdeI and BamHI, and subcloned into NdeI and BamHI digested pZ1B. This digestion removed the FGF and 5' portion of SAP (up to the BamHI site) from the parental FGF-SAP vector (pZ1B) and replaced this portion with a SAP molecule containing a CYS at position +4 or +10 relative to the start site of the native mature SAP protein (see SEQ ID NOs. 36 and 37, respectively). The resulting plasmids are designated pZ51B and pZ52B, respectively.

D. Cloning of DNA encoding SAP mutants in vector pET15b

1. The SAP-Cys-1 mutants

The initial step in this construction was the mutagenesis of the internal BamHI site at nucleotides 555-560 (SEQ ID NO. 34) in pZ1B using a sense primer corresponding to nucleotides 543-570 (SEQ ID NO. 34) but changing the G at nucleotide 555 (the third position in the Lys codon) to an A. The complement of the sense primer was used as the antisense primer: 5' TTTCAGGTTTGGATCTTTTACGTTGTTT 3' (SEQ ID NO. 61). The first round of amplification used amplification reactions, conducted as in B above, with primers having SEQ ID NOs. 15 and 61 (set forth above) and primers having SEQ ID NOs. 62 and 24 as follows:

5' AAACAACGTAAAAGATCCAAACCTGAAA 3' (SEQ ID NO. 62)
5' GGATCCGCCTCGTTTGACTACTT 3' (SEQ ID NO. 24).

Individual fragments were gel purified and a second round of amplification was performed using primers of SEQ ID Nos. 15 and 24, performed as in B., above. This amplification introduced a NdeI site and a Cys codon onto the 5' end of the saporin-encoding DNA. The antisense primer was complementary to the 3' end of the saporin protein and encoded a BamHI site for cloning and a stop codon (SEQ ID NO. 24):

The resulting fragment was digested with *Nde*I/*Bam*HI and inserted into pET15b (Novagen, Madison, WI), which has a His-Tag™ leader sequence (SEQ ID NO. 23), that had also been digested with *Nde*I/*Bam*HI. The sequence of SAP-Cys-1 is set forth in SEQ ID NO. 58).

2. The SAP-Cys+4 and Sap-Cys+10 mutants

This construction was performed similarly to the SAP-Cys-1 using pZ1B as the starting material, and splice overlap extension (SOE) using PZ1B as the starting plasmid, including mutagenesis of the internal BamHI site at nucleotides 555-560 (SEQ ID NO. 34) in pZ1B using a sense primer corresponding to nucleotides 543-570 (SEQ ID NO. 34) but changing the G at nucleotide 555 (the third position in the Lys codon) to an A and introduction of the cys at position 4 or 10 in place of the native amino acid. The first round of amplification used primers of SEQ ID NOs. 59 and 61 (for the cys+4 saporin mutants) or SEQ ID NOs. 60 and 61 for the cys+10 saporin mutants): CATATGGTCACATCATGTACATTAGATCTAGTAAAT (SEQ ID NO. 59); and CATATGGTCACATCAATCACATTAGATCTAGTATGTCCGACCGCGGGTCA (SEQ ID NO. 60); TTTCAGGTTTGGATCTTTTACGTTGTTT (SEQ ID NO. 61). For each construction, the second round of amplification included the fragment prepared in D.1., above, using primers having SEQ ID NOs. 62 and 24.

Amplification conditions were as follows: denaturation for 1 min at 94° C, annealing for 2 min at 70° C and extension for 2 min at 72° C for 35 cycles. Individual fragments were gel purified and subjected to a second round of amplification, following the same protocol, using only the external oligonucleotides of SEQ ID NO. 24 and SEQ ID NO. 59 for the cys+4 mutant or SEQ ID NOs. 60 and 24 for the cys+10 mutant. The resulting fragments had a *Nde*I site on the 5' end of the saporin-encoding DNA and a BamHI site for cloning and a stop codon on the 3' end. The resulting fragment was digested with *Nde*I/*Bam*HI and inserted into pET 15b (Novagen, Madison, WI), which has a His-Tag™ leader sequence (SEQ ID NO. 23), that had also been digested NdeI/*Bam*HI.

DNA encoding unmodified SAP (EXAMPLE 1) can be similarly inserted into a pET15b or pET11A and expressed as described below for the modified SAP-encoding DNA.

E. Expression of the modified saporin-encoding DNA

BL21(DE3) cells were transformed with the resulting plasmids and cultured as described in Example 1, except that all incubations were conducted at 30° C instead of 37° C. Briefly, a single colony was grown in LB AMP_{100} to an OD_{600} of 1.0-1.5 and then induced with IPTG (final concentration 0.2 mM) for 2 h. The bacteria were spun down.

F. Purification of modified saporin

Lysis buffer (20 mM $NaPO_4$, pH 7.0, 5 mM EDTA, 5 mM EGTA, 1 mM DTT, 0.5 µg/ml leupeptin, 1 µg/ml aprotinin, 0.7 µg/ml pepstatin) was added to the rSAP cell paste (produced from pZ50B in BL21(DE3) cells, as described above) in a ratio of 1.5 ml buffer/g cells. This mixture was evenly suspended via a Polytron homogenizer and passed through a microfluidizer twice.

The resulting lysate was centrifuged at 50,000 rpm for 45 min. The supernatant was diluted with SP Buffer A (20 mM $NaPO_4$, 1 mM EDTA, pH 7.0) so that the conductivity was below 2.5 mS/cm. The diluted lysate supernatant was then loaded onto a SP-Sepharose column, and a linear gradient of 0 to 30% SP Buffer B (1 M NaCl, 20 mM $NaPO_4$, 1 mM EDTA, pH 7.0) in SP Buffer A with a total of 6 column volumes was applied. Fractions containing rSAP were combined and the resulting rSAP had a purity of greater than 90%.

A buffer exchange step was used to get the eluate into a buffer containing 50 mM $NaBO_3$, 1 mM EDTA, pH 8.5 (S Buffer A). This sample was then applied to a Resource S column (Pharmacia, Sweden) pre-equilibrated with S Buffer A. Pure rSAP was eluted off the column by 10 column volumes of a linear gradient of 0 to 300 mM NaCl in SP Buffer A. The final rSAP was approximately 98% pure and the overall yield of rSAP was about 50% (the amount of rSAP in crude lysate was determined by ELISA).

In this preparation, ultracentrifugation was used to clarify the lysate; other methods, such as filtration and using floculents also can be used. In addition, Streamline S (PHARMACIA, Sweden) may also be used for large scale preparations.

EXAMPLE 4

CONSTRUCTION OF PLASMIDS ENCODING HBEGF-SAP FUSION PROTEINS

A. Materials

1. Bacterial Strains and Plasmids

E. coli strains BL21(DE3), BL21(DE3)pLysS, HMS174(DE3) and HMS174(DE3)pLysS were purchased from Novagen, Madison, WI.

2. DNA Manipulations

The restriction and modification enzymes employed here are commercially available in the U.S. Minipreparation and maxipreparations of plasmids, preparation of competent cells, transformation, M13 manipulation, bacterial media and Western blotting were performed using routine methods (*see, e.g.*, Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). The purification of DNA fragments was done using the Geneclean II kit, purchased from Bio 101. SDS gel electrophoresis was performed on a Phastsystem (Pharmacia).

B. Removal of FGF Sequences from PZ1B1

The plasmid PZ1B1 contains DNA encoding FGF linked to DNA encoding saporin via a spacer region encoding two amino acids (Ala-Met). The fusion gene is cloned into the *Nde*I and *Bam*HI sites of the plasmid vector pET11a (Novagen). The vector provides a T7lac promoter, a lac operator, and a ribosome binding site upstream of the fusion gene, and a T7 terminator downstream of the fusion gene.

To remove FGF sequences, PZ1B1 was digested with the restriction enzymes *Nde*I and *Nco*I. *Nde*I cuts once within the plasmid at a position encompassing the translation initiation codon (ATG) for FGF. *Nco*I also cuts once within the plasmid at a site within the two amino acid linker region of the plasmid PZ1B1. Digestion of PZ1B1 with *Nde*I and *Nco*I thus generates two fragments: a FGF-fragment and a fragment containing vector (pET11a) and saporin-encoding sequence. The digestion products were resolved in an agarose gel and the vector/saporin fragment was purified using the Geneclean II kit.

C. Amplification and isolation of DNA encoding mature HBEGF

The plasmid pJMU2-1 (see, gift from Dr. J. Abraham; see, also International PCT Application WO 92/06705, which is based on U.S. Application Serial No. 08/598.082) contains a 2.36 kb human HBEGF cDNA fragment (with added *Eco*RI linkers) cloned into the *Eco*RI site of pUC9 (*see, e.g.*, Viera et al. (1982) *Gene 19*:259-2678; see, also GB 2045254 A; available from numerous commercial sources, such as

Pharmacia Fine Chemicals, Piscataway, NJ). This fragment encodes a 208 amino acid precursor form of HBEGF (SEQ ID NO. 1). Any plasmid that contains the cDNA encoding full length precursor protein (e.g., pMTN-HBEGF, ATCC #40900 and pAX-HBEGF, ATCC #40899) may also be used for HBEGF amplification as described herein.

The region of the DNA that encodes a 77 amino acid form of mature HBEGF (corresponding to nucleotides 217-447 or amino acids 73-149 in the precursor) was amplified. The primer (SEQ ID NO. 25) corresponding to the "sense" strand of mature HBEGF includes an NdeI restriction site adaptor (and a Met codon) upstream of the codon for amino acid 73 of precursor HBEGF (SEQ ID No. 1) and spans the first 14 nucleotides of the DNA sequence encoding mature HBEGF (SEQ ID NO. 3):

5'-CTGGACCATATGAGAGTCACTTTA-3' (SEQ ID NO. 25).

The primer (SEQ ID NO. 26) corresponding to the "antisense" strand of HBEGF complements 23 nucleotides encoding amino acids 144-149 of the precursor peptide and introduces an *Rca*I restriction site downstream of the HBEGF-encoding DNA (SEQ NO. 26):

5'-GTATATCATGACTGGGAGGCTCAGCCCATGACA-3'

An amplification reaction was performed in which plasmid pJMU2-1 DNA (200 ng) was mixed in a final volume of 100 µl containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 µM dNTPs, 100 pmole of each primer, and 2.5 U TaqI polymerase (Boehringer Mannheim). Amplifications were done in a TwinBlock DNA thermal cycler (Ericomp). The first cycle was a denaturation step (94°C for 5 min). The second cycle was repeated 30 times and included a denaturation step, an annealing step, and an elongation step (94°C for 1 min; 62°C for 2 min; 72 °C for 2 min). The third cycle was an elongation step (72 °C for 7 min). An aliquot of the reaction was run on an agarose gel and the amplified product was purified using the Geneclean II kit (Bio 101).

D. Preparation of plasmids that encode the HBEGF-SAP fusion protein

The purified amplified product encoding HBEGF was then digested with *Nde*I and *Rca*I (which generates an end compatible with *Nco*I) and ligated into the *Nde*I/*Nco*I sites of PZ1B1. Following transformation into the E. Coli strain NovaBlue (Novagen) positive clones were identified by restriction enzyme digestion of miniprep

DNAs. A positive clone designated pZ31B1 was sequenced starting within the vector sequence (using the T7 promoter primer) and extending through the HBEGF-coding sequence, the Val-Met two amino acid linker (generated by ligation at the *Rca*I and *Nco*I sites), and into the saporin sequence. The positive pZ31B1 plasmid gave the proper nucleotide sequence (*i.e.*, SEQ ID NO. 6) for the HBEGF-SAP fusion gene. The fusion protein encoded by the plasmid pZ31B1 contains 78 amino acids of HB-EGF (including a methionine introduced by the *Nde*I restriction site), a two amino acid (Val-Met) linker and 253 amino acids of saporin (SEQ ID NO. 6). The pZ31B1 encoded fusion protein is therefore 333 amino acids long with a predicted molecular weight of about 37.6 kD and an isoelectric point of 9.6.

E. Expression of the recombinant HBEGF-SAP fusion proteins

The two-stage method described above was used to produce recombinant HBEGF-SAP protein (hereinafter HBEGF-SAP fusion protein) encoded by pZ31B1. Plasmid pZ31B1 was transformed into *E. coli* strain BL21(DE3), which contains chromosomal copies of the T7 RNA polymerase gene linked to an IPTG-inducible lacUV promoter.

1. Small scall preparation

The transformed *E. coli* cells were grown in 50 ml cultures of LB broth + ampicillin (100 μg/ml) at 30°C to an OD_{600} of 0.7. The second stage was commenced by the addition of IPTG (0.1 mM) to induce expression of the T7 RNA polymerase gene. Cultures were continued at 30° C. One ml aliquots of the culture were removed just prior to IPTG addition and then hourly thereafter. Aliquots were centrifuged, resuspended in 1 ml lysis buffer (10 mM Tris pH 8.0, 2 mM EDTA, 0.01 mg/ml lysozyme, 10 mM DTT) and incubated for 1 hour at room temperature. Following centrifugation, the lysed supernatants were analyzed by Western blotting (using an anti-SAP antibody) for expression of the fusion protein. SDS gel electrophoresis was performed on a Phastsystem utilizing 10-15% gradient gels (Pharmacia). Western blotting was accomplished by transfer of the electrophoresed protein to nitrocellulose using the PhastTransfer system (Pharmacia), as described by the manufacturer. Anti-SAP antibodies are used at a dilution of 1:1000. Horseradish peroxidase labeled anti-IgG was used as the second antibody (Davis et al. (1986) Basic Methods in Molecular Biology, New York, Elsevier Science Publishing Co., pp 1-338). The Western blot analysis demonstrated induction of a soluble protein with the predicted molecular weight.

Aliquots of the bacterial lysates were also analyzed by an ELISA assay (using an anti-SAP antibody). The results of this assay confirmed the induced expression of HBEGF-SAP.

2. Large-scale preparation

Three liters of IPTG-induced bacterial culture were grown as described above, in a fermentation apparatus except that carbenicillin (100 μg/ml) was substituted for ampicillin. The pelleted culture was stored at -80°C.

F. Purification of HBEGF-SAP fusion protein

Aliquots of the fermentation culture paste were removed from the freezer, resuspended in Buffer A (10 mM NaCitrate -pH 6.0, 10 mM EDTA, 10 mM EGTA, 50 mM NaCl), and lysed with a Microfluidizer (Model 110Y, Microfluidics Corp.). The lysate was centrifuged at 100,000 x g and the resulting supernatant was loaded onto a S-Streamline column (Pharmacia) equilibrated with buffer A. The column was washed with buffer B (10 mM Na-Phosphate, 5 mM EDTA, 5 mM EGTA at pH 8.0) containing 0.2 M NaCl until the A_{280} of the eluate reached baseline. The HBEGF fusion protein was eluted with buffer B containing 0.8 M NaCl.

Fractions were analyzed for presence of the fusion protein by SDS PAGE and Western blotting. The HBEGF-containing fractions were pooled, diluted 4x with buffer B and applied to a Q-Sepharose (Pharmacia) column equilibrated with buffer B. The flow through was applied directly to a SP-Sepharose Fast Flow cation exchange column (Pharmacia) equilibrated with buffer B containing 0.2 M NaCl. The HBEGF fusion protein was eluted with a 0.2-1.0 M NaCl gradient. Fractions containing fusion protein (as determined by SDS PAGE) were pooled and loaded onto a heparin-Sepharose CL6B (Pharmacia) column equilibrated with buffer C (10 mM NaCitrate-pH 6.0, 1 mM EDTA, 1 mM EGTA, 0.2 M NaCl). The fusion protein was eluted with a 0.2-1.2 M NaCl gradient. Fractions containing fusion protein were pooled and NH_4SO_4 was added to 2.0 M. Following filtration, the material was applied to a Phenyl-Sepharose HP column equilibrated with buffer C containing 2 M NH_4SO_4. The fusion protein was eluted with a 2.0-0.0 M NH_4SO_4 gradient. Fractions containing fusion protein were pooled and applied to a S100 size exclusion column equilibrated with 10 mM NaCitrate (pH 6.0), 0.1 M EDTA, 0.14 M NaCl. Fractions containing purified HBEGF fusion protein were then selected.

G. Characterization of the HBEGF-SAP fusion protein

1. Effect of HBEGF-SAP fusion protein on cell-free protein synthesis

The RIP activity of HBEGF fusion protein encoded by pZ31B1 was assayed as described in Example 1 for saporin. The results indicated that the IC_{50} of the HBEGF-SAP fusion protein exhibits activity in this assay.

2. Cytotoxicity of HBEGF-SAP fusion protein

Cytotoxicity experiments are performed with the Promega (Madison, WI) CellTiter 96 Cell Proliferation/Cytotoxicity Assay. About 1,500 A431 cells (ATCC Accession No. CRL 1555), an epidermoid carcinoma cell line, are plated per well in a 96 well plate in 90 µl HDMEM plus 10% FCS and incubated overnight at 37° C, 5% CO_2. The following morning 10 µl of media alone or 10 µl of media containing various concentrations of the fusion protein, HBEGF polypeptide or saporin are added to the wells. The plate is incubated for 72 hours at 37 C. Following the incubation period, the number of living cells is determined by measuring the incorporation and conversion of the commonly available dye MTT supplied as a part of the Promega kit. Fifteen µl of the MTT solution is added to each well, and incubation is continued for approximately 4 hours. Next, 100 µl of the standard solubilization solution supplied as a part of the Promega kit is added to each well. The plate is allowed to stand overnight at room temperature and the absorbance at 560 nm is read on an ELISA plate reader (e.g., Titertek Multiskan PLUS, ICN, Flow, Costa Mesa, CA).

EXAMPLE 5

CHEMICAL SYNTHESIS OF HBEGF-SAP

About 50-100 nmol of HBEGF that has been dialyzed against phosphate-buffered saline is added to about 2.5 mg mono-derivatized SAP (a 1.5 molar excess over the HBEGF polypeptide) and left on a rocker platform overnight. The ultraviolet-visible wavelength spectrum is checked in order to determine the extent of reaction by the release of pyridylthione, which adsorbs at 343 nm with a known extinction coefficient. The reaction mixtures are treated for purification in the following manner: reaction mixture is passed over a HiTrap heparin-Sepharose column (1 ml) equilibrated with 0.15 M sodium chloride in buffer A at a flow rate of 0.5 ml/min. The column is washed with 0.6 M NaCl and 1.0 M NaCl in buffer A and the product eluted with 4.0 M NaCl in buffer A. Fractions (0.5 ml) are analyzed by gel electrophoresis and

absorbance at 280 nm. Peak tubes are pooled and dialyzed versus 10 mM sodium phosphate, pH 7.5 and applied to a Mono S 5/5 column equilibrated with the same buffer. A 10 ml gradient between 0 and 1.0 M sodium chloride in equilibration buffer is used to elute the product.

Cytotoxicity of HBEGF-SAP

Cytotoxicity to several cell types, such as A-431 cells (ATCC Accession No. CRL 1555) or other smooth muscle cells is tested using the Promega (Madison, WI) CellTiter 96 Cell Proliferation/Cytotoxicity Assay described above in Example 4. The HBEGF-SAP conjugate should be cytotoxic to each cell type that expresses an EGF receptor.

EXAMPLE 6

CONSTRUCTION OF PLASMIDS FOR INSERTION OF LINKERS

A. Construction of plasmid PZ32B1 containing linker-amenable HBEGF-SAP by mutation of *Nco*I sites within the coding region of mature HBEGF

Plasmid plasmid pJMU2-1, as described in Example 4, was used as the amplification template for preparatin of the linker-amenable HBEGF-SAP plasmid pZ32B1. Each of the two *Nco*I sites contained within the region encoding mature HBEGF was mutated in separate amplification reactions. First, a "sense" primer was constructed that corresponds to the nucleotides encoding amino acids 13-19 (SEQ ID NO. 1, nucleotides 37-57 in the HBEGF precursor and includes a PstI site (SEQ ID NO. 51):

*Pst*I

5'-CTGGCTGCAGTTCTCTCGGCA-3'.

An "antisense" primer spanning the nucleotides that encode amino acids 114 to 129 in the HBEGF precursor was designed that introduces a single base mutation (T→C in the sense strand) that destroys an NcoI site while maintaining a codon for the amino acid histidine at position 118 (SEQ ID NO. 29):

SacI

5'-AGCCCGGAGCTCCTTCACATATTTGCATTCTCCGTGGATGCAGAAG-3'.

The antisense primer also spans a *Sac*I site. Amplification of the DNA encoding HBEGF using these two primers generates a fragment with a *Pst*I site at its 5' end and a *Sac*I site at the 3' end. The amplification was carried out under the same conditions previously described in Example 4.C., except that the second cycle was repeated only 20 times.

The second *Nco*I site within the mature HBEGF coding sequence was mutated using a "sense" primer that spans the nucleotides encoding amino acids 124-142 in the HBEGF precursor (SEQ ID NO. 30):

SacI
5-GTGAAGGAGCTCCGGGCTCCCTCCTGCATCTGCCACCCGGGTTATCATGGAGAGAGG-3'

This sense primer includes a *Sac*I site, and introduces a single base mutation (C→T) that destroys the *Nco*I site while maintaining a codon for the amino acid tyrosine at position 138 of SEQ ID NO. 1.

An antisense primer that spans a region of the HBEGF-encoding DNA downstream of the precursor HBEGF coding region was designed to introduce an *Eco*RI site adaptor at the 3' end of the amplified DNA fragment (SEQ ID NO. 31):

EcoRI
5'-ATATAGAATTCTGTCTTCTCAGAGGTA-3'.

Amplification of HBEGF-encoding DNA using these two primers generates a fragment with a *Sac*I site at its 5' end and an *Eco*RI site at the 3' end.

The amplified HBEGF fragments generated by the two above amplification reactions overlap at a *Sac*I site. Following purification using the Geneclean II kit (Bio 101), the first product was digested with *Pst*I and *Sac*I and the second product was digested with *Sac*I and *Eco*RI. The digested fragments were ligated into the *Pst*I and *Eco*RI sites of the vector pGEM-4 (the pGEM series of plasmids are available from Promega, Madison WI; *see also*, U.S. Patent No. 4,766,072, which describes construction of the pGEM plasmids) producing the plasmid pGEM/HBEGF. This plasmid contains a regenerated colinear piece of DNA encompassing the entire mature HBEGF coding region (*see, e.g.*, nucleotides 1-234 of SEQ ID No. 33).

Using this pGEM/HBEGF plasmid as a template, the mature HBEGF encoding region was amplified using the primers set forth in SEQ ID NO. 25 that corresponds to the "sense" strand of mature HBEGF including an *Nde*I restriction site adaptor just upstream of the codon for amino acid 73 of precursor HBEGF. The other

primer corresponds to the "antisense" strand of HBEGF spanning nucleotides encoding amino acids 143-149 of precursor HBEGF (*see, e.g.*, SEQ. ID NO. 1), and introduces an Ala-Met-NcoI restriction site just downstream of mature HBEGF-encoding DNA (SEQ ID NO. 32):

NcoI

5'- ATATACCATGGCTGGGAGGCTCAGCCCATGACA-3'

Amplification of HBEGF sequences using these two primers and the above pGEM/HBEGF plasmid as a template, generates a mature HBEGF encoding fragment with an NdeI site at the 5' end and a unique *Nco*I site at the 3' end. An aliquot of the amplification reaction was run on an agarose gel and the amplified product was purified using the Geneclean II kit. The purified DNA was then digested with *Nde*I and *Nco*I and ligated into the NdeI/NcoI sites of pZ1B1 (digested to remove FGF-encoding DNA. as described in Example 4.B.).

Following transformation into the *E. coli* strain NovaBlue (Novagen) positive clones were identified by restriction enzyme digestion of miniprep DNA. A positive clone designated pZ32B1 plasmid was sequenced starting within the vector sequence (using the T7 promoter primer) and extending through the HBEGF-coding sequence, the two amino acid Ala-Met linker and into the saporin sequence. The plasmid pZ32B1 gave the desired sequence for the HBEGF-SAP fusion gene set forth in SEQ ID NO. 33. The fusion protein encoded by the plasmid pZ32B1 includes 78 amino acids of HBEGF (including a methionine introduced by the *Nde*I restriction site), a two amino acid (Ala-Met) linker and 253 amino acids of saporin (SEQ ID NO. 33). The fusion protein is therefore 333 amino acids long with a predicted molecular weight of about 37.6 kD and an isoelectric point of 9.6.

The resulting linker-amenable HBEGF-SAP plasmid pZ32B1 differs from the HBEGF-SAP encoding plasmid PZ31B1 described in Example 4.D. in the following ways:

1) Two *Nco*I sites within the coding region for mature HBEGF have been mutated (by amplification) so that the *Nco*I sites are destroyed without changing the reading frame or amino acid composition of HBEGF.

2) The two amino acid linker between HBEGF and SAP is Ala-Met in the new plasmid construct pZ32B1. This Ala-Met linker encompasses an *Nco*I site (the only *Nco*I site in the new plasmid). Therefore, the resulting HBEGF-SAP plasmid can be linearized by digestion with *Nco*I. Different linkers, which have *Nco*I sites at their ends, can then be inserted between the HBEGF and SAP sequences.

The desired linker is then inserted into plasmid pZ32B1. The resulting plasmid is introduced into *E. coli* host cells, expressed and the fusion proteins isolated as described in Example 4.F. Fusion proteins may also be isolated using the same procedures as those described for HBEGF (*see, e.g.*, International PCT Application WO 92/06705, which is based on U.S. Application Serial No. 08/598,082).

B. Preparation of a PETSAP-MCS (MCS=multiple cloning sites)

A SAP cassette plasmid (PETSAP-MCS) was made that would be amenable to insertion of any growth factor sequence downstream from the saporin-encoding DNA.

SAP encoding DNA was amplified (using PZ32B1 as a template) to give SEQ ID NO. 81. The sense and antisense primers respectively, used to amplify this SAP fragment were:

SEQ ID NO. 54: 5'-TGAGCGAATTCCATATGGTCACATCAATCACATTA
EcoRI NdeI

SEQ ID NO. 55: 5'TATATGAATTCCATGGCCTTTGGTTTGCCCAAATACAT
EcoRI NcoI

The resulting SAP-encoding DNA has an *Eco*RI site at its 5' end followed by an *Nde*I site that encompasses the ATG codon. The 3' end of the SAP fragment has no stop codon, and has an *Nco*I site followed by an *Eco*RI restriction site.

The amplified SAP fragment was then digested with *Eco*RI and subcloned into the *Eco*RI site of the plasmid pGEM-4 (pGEM-4 serves as the source of the MCS, the pGEM series of plasmids are available from Promega, Madison WI; *see also*, U.S. Patent No. 4,766,072, which describes construction of the pGEM plasmids) in such an orientation that the multicloning site (MCS) of pGEM-4 lies downstream (3' of) from the SAP -encoding DNA.

Plasmid pGEMSAP was digested with *Pst*I and the ends of the fragment were blunt-ended. The fragment was then digested with *Nde*I, thereby generating a fragment that contains all of the saporin-encoding DNA and most of the MCS of pGEM-4. This fragment was then cloned into the *Nde*I/*Bam*HI sites of pET 11a, in which the *Bam*HI site had been blunt-ended by filling in with Klenow polymerase. The resulting plasmid was designated PETSAP-MCS. It has unique *Sac*I, *Sma*I, and *Sal*I sites in the MCS for insertion of DNA encoding a desired linker, HBEGF, or combination of HBEGF and linker downstream from (3' of) the saporin-encoding DNA.

EXAMPLE 7

PREPARATION OF HBEGF CONJUGATES THAT CONTAIN LINKERS

A. Synthesis of oligonucleotides encoding protease substrates and oligonucleotides encoding flexible linkers

Complementary single-stranded oligos in which the sense strand encodes a protease substrate or flexible linker, have been synthesized either using a Cyclone machine (MILLIPORE, Bedford, MA) according the instructions provided by the manufacturer or, if greater than 80 bases, were made by Midland Certified Reagent Co. (MIDLAND, TX). The following oligos have been synthesized and can be introduced into constructs encoding HBEGF-SAP or SAP-HBEGF.

1. Cathepsin B substrate linker:

5'- CCATGGCCCTGGCCCTGGCCCTGGCCCTGGCCATGG SEQ ID NO. 38

2. Cathepsin D substrate linker

5'- CCATGGGCCGATCGGGCTTCCTGGGCTTCGGCTTCCTGG
GCTTCGCCAT GG -3' SEQ ID NO. 39

3. Trypsin substrate linker

5'- CCATGGGCCGATCGGGCGGTGGGTGCGCTGGTAATAGAGT
CAGAAGATCAGTCGGAAGCAGCCTGTCTTGCGGTGGTCTC
GACCTGCAGG CCATGG-3' SEQ ID NO. 44

4. Gly_4Ser

5'- CCATGGGCGG CGGCGGCTCT GCCATGG -3' SEQ ID NO. 40

5. $(Gly_4Ser)_2$

5'- CCATGGGCGGCGGCGGCTCTGGCGGCGGCGGCTC
TGCCATGG -3' SEQ ID NO. 41

6. $(Ser_4Gly)_4$

5'- CCATGGCCTCGTCGTCGTCGGGCTCGTCGTCGTC
GGGCTCGTCGTCGTCGGGCTCGTCGTCGTCGGGC
GCCATGG -3' SEQ ID NO. 42

7. $(Ser_4Gly)_2$

5- CCATGGCCTCGTCGTCGTCGGGCTCGTCGTCGTC
GGGCGCCATGG -3' SEQ ID NO. 43

8. Thrombin substrate linker

CTG GTG CCG CGC GGC AGC SEQ ID NO. 45
Leu Val Pro Arg Gly Ser

9. Enterokinase substrate linker

GAC GAC GAC GAC CCA SEQ ID No. 46
Asp Asp Asp Asp Lys

10. Factor Xa substrate

ATC GAA GGT CGT SEQ ID No. 47
Ile Glu Gly Arg

B. Preparation of DNA constructs encoding HBEGF-Linker-SAP

HBEGF-Ala-Met-SAP (PZ32B1) was digested with NcoI for insertion of linker sequences. The following linkers have been inserted: cathepsin D sensitive site, diphtheria toxin trypsin sensitive site, and $Gly_4SerGly_2SerGly_4SerGly_4Ser$, which may enhance binding of the fusion protein to the receptor compared to fusion proteins lacking such linker.

C. Preparation of SAP-Ala-Met-Ala-HBEGF

HBEGF-encoding DNA was amplified (using PZ32B1 as a template) to produce a NcoI site at the 5' end and a stop codon followed by a SalI site at the 3' end. The sense and antisense primers, respectively, used in the amplification reaction were:

SEQ ID No. 52: 5'-TATATGCCATGGCCAGAGTCACTTTATCCTCCAAG
NcoI

SEQ ID No. 53: 5'-TATATGTCGACTATGGGAGGCTCAGCCCATGACA
SalI stop

The resulting amplified product was digested with NcoI and SalI and ligated into NcoI/SalI digested PETSAP-MCS. The resulting plasmid (PZ36B1) encodes a protein with an Ala-Met-Ala linker between the SAP and HBEGF moieties (SEQ ID No. 82).

D. Preparation of SAP-Ala-Met-$(Gly_4Ser)_4$-Ala-Met-Ala-HBEGF

Plasmid PZ36B1 was digested with NcoI and a linker encoding Ala-Met-$(Gly_4Ser)_4$-Ala-Met-Ala was inserted. The resulting plasmid was designated PZ37B1.

E. Expression of conjugates with linkers

DNA encoding the conjugates set forth above and summarized in Table 3 are expressed above for PZ31B1 using plasmids prepared as described above and summarized in TABLE 3.

F. Western blot analysis of HBEGF fusion proteins

All HBEGF constructs have demonstrated inducible expression of proteins of the expected size when analyzed by Western blotting. Using the protocol set forth in Example 4. F., purification (to greater than 95%) of HBEGF-Ala-Met-

$Gly_4SerGly_2Ser$ $(Gly_4Ser)_2$-Ala-Met-SAP and SAP-Ala-Met-$(Gly_4Ser)_{x4}$-Ala-Met-Ala-HBEGF has been achieved. Specifically, when purification of HBEGF-Ala-Met-$(Gly_4SerGly_2Ser)(Gly_4Ser)_2$-Ala-Met-SAP was optimized, immunoreactive material eluted from the heparin sepharose column in two peaks, the first peak eluting at 0.6-0.8 M NaCl (pool A) and the second peak eluting at 0.9-1.0 M NaCl (pool B). Pool B was found to contain material whose bioactivity (on A431 cells) was ten times more active than the material in Pool A.

G. Biological activity of HBEGF fusion proteins

The fusion protein HBEGF-Val-Met-SAP (encoded by plasmid PZ31B1) was active in the cell-free RIP assay

Insertion of the $(Gly_4Ser)_{x4}$ linker into HBEGF-SAP has generated a fusion protein with cytotoxicity to A431 cells (ID_{50} on the order of 10^{-10} -10^{-9} M). The purified fusion protein SAP-Ala-Met-$(Gly_4Ser)_{x4}$-Ala-Met-Ala-HBEGF exhibits similar, perhaps somewhat higher, cytotoxicity. These two HBEGF fusion proteins have also been tested for their cytotoxicity (relative to FGF-SAP) using other cell lines including aortic smooth muscle cells (active), glioblastoma and medulloblastoma cells (active), SK-MEL melanoma cells (somewhat active), and small cell lung carcinoma cells (inactive). Therefore, there are cell-type differences in the cytotoxicity of these proteins.

EXAMPLE 8

BACULOVIRUS EXPRESSION OF HBEGF

The following proteins have been expressed in the baculovirus system: Met-Cys-HBEGF, Met-Cys-Ala-Met-Ala-HBEGF (linker amenable), Met-Cys-Ala-Met-$(Gly_4Ser)_2$-Ala-Met-Ala-HBEGF (prepared by insertion of linker into the NcoI site of Met-Cys-Ala-Met-Ala-HBEGF), and Met-Cys-Ala-Met-$(Gly_4Ser)_{x4}$-Ala-Met-Ala-HBEGF (prepared by insertion of linker into the NcoI site of Met-Cys-Ala-Met-Ala-HBEGF).

A. Modification of HBEGF encoding DNA

Mature HBEGF-encoding DNA fragments were amplified (using PZ32B1 as a template) to give a BamHI site at the 5' end followed by Met-Cys codons. At the 3' end the amplified product had a stop codon followed by a *Hind*III site. The primers used were (SEQ ID NOs. 56 and 57, respectively):

Sense: 5'-TATATGGATCCTATGTGTAGAGTCACTTTATCCTCCAAG
BamH Met Cys
Antisense: 5'-TATATAAGCTTCTATGGGAGGCTCAGCCCATGACA
HindIII STOP

The amplified product was digested with BamHI and HindIII and ligated into BamHI/*Hind*III digested pBlueBacIII (Invitrogen). The sequence of the DNA encoding HBEGF in this plasmid is set forth in SEQ ID NO. 83.

B. Preparation of a linker-amenable HBEGF-baculovirus vector

A linker-amenable HBEGF/BlueBac clone was made by amplifying HBEGF sequences as above using a different sense primer (SEQ ID NO. 86):

5' TATAGGATCCTGATGTGTGCCATGGCCAGAGTCACTTTATCCTCCAAGCCA
BamHI Met Cys Ala Met Ala

The resulting amplified fragment (SEQ ID No. 84) has a BamHI site at the 5' end followed by Met-Cys-Ala-Met-Ala (SEQ ID NO. 85) codons that encompass a NcoI site.

EXAMPLE 9

IN VIVO ASSAYS FOR MONITORING THE EFFECTS OF CONJUGATES ON SMOOTH MUSCLE CELLS

In vivo assays monitoring the effects of conjugates on smooth muscle cells have been described, for example, in Casscells et al. (1992) *Proc. Natl. Acad. Sci. USA 89*:7159-7163. Such assays are used herein.

Balloon catheter denudation is performed on the left carotid artery of 5-6 month old male Sprague-Dawley rats by intraluminal passage of a 2F Fogarty balloon. Body weights range from 300-350 g the day prior to surgery. At 0, 3, 6, 9 days after balloon injury, wild-type chemical conjugate HBEGF-SAP (1-10 μg/kg/dose), fusion protein HBEGF-SAP (1-10 μg/kg/dose), or vehicle (0.9% NaCl, 0.1% human serum albumin (HSA)) is injected via the tail vein. The therapeutic composition is prepared by mixing the test materials with appropriate volumes of 0.9% NaCl, 0.1% HSA. The wild-type chemical conjugate is supplied in 10 mM citrate, 0.14 m NaCl.

0.1 mM EDTA, pH 6 at a concentration of 1.0 mg/ml prior to being prepared in appropriate dosages. The fusion protein is supplied in 10 mM citrate, 0.14 m NaCl, 0.1 mM EDTA, pH 6 at a concentration of 0.256 mg/ml prior to being prepared in appropriate dosages. On day 14 after balloon denudation, approximately 120 hr after the final dose, animals are sacrificed with an overdose of intravenous KCl under deep anesthesia. One hour before sacrifice, animals are injected intravenously with Evans blue dye (0.5 ml, 5% in saline) to confirm endothelial denudation. At one and 17 hours prior to sacrifice, animals are injected intraperitoneally with Bromodeoxyuridine (BrdU, 30 mg/kg) for quantitation of cellular proliferation. At sacrifice, the arterial tree is perfused at 80 mm Hg with Hank's balanced salt solution, 15 mM HEPES, pH 7.4 until the perfusate from the jugular is clear of blood. The arterial tree is then perfused with 2% paraformaldehyde in 0.1 M Na Cacodylate buffer, pH 7.4, for 15 minutes. The carotid arteries are then removed, cut into sections, and processed for light microscopy. Tissue samples are dehydrated and embedded in paraffin, cut into 4μ sections, and stained with hematoxylin-eosin and Movat pentachrome stain. Vessels are then measured for intimal, medial, and neointimal areas by computerized planimetry. Anti-BrdU antibody is used for detection of BrdU positive cells; smooth muscle cell proliferation is quantitated by counting BrdU positive cells as a percent of total smooth muscle cells.

EXAMPLE 10

EFFECT OF HBEGF-CONTAINING CONJUGATE IN MOUSE SOLID TUMOR XENOGRAFT MODEL

The *in vivo* mouse solid tumor xenograft model, which assays for a compound's ability to inhibit tumor cell proliferation, has been described, in Beitz et al. (1992) *Cancer Res. 52*:227-230. For example, wild-type chemical conjugate and fusion protein HBEGF-SAP are evaluated for anti-tumor activity against any EGF-receptor expressing tumor subtype, e.g., bladder carcinoma, in a mouse tumor xenograft model. Sixty-three athymic nude mice (25 to 30 g) bearing subcutaneous tumors are randomized into nine treatment groups (n=7/treatment) and given four weekly bolus intravenous injections of wild-type chemical conjugate HBEGF-SAP (0.5 μg/kg and 50 μg/kg), fusion protein HBEGF -SAP (0.5, 5.0, and 50 μg/kg), SAP only (85 μg/kg), HBEGF only (50 μg/kg), SAP with HBEGF (85 and 50 μg/kg, respectively), or vehicle

(PBS with 0.1% BSA). Dosing material is prepared by mixing the test material with appropriate volumes of PBS/0.1% BSA to achieve the desired doses. Individual syringes are prepared for each animal. Mice receive four weekly IV injections (250-300 ul) into the tail vein on days 5, 12, 19 and 26 with day 1 designated as the day that the tumor cells are injected into the mice. Doses are individualized for differences in body weight. Tumor volume is measured twice weekly for a period of 61 days.

Female Balb/c nu/nu athymic mice (Roger Williams Hospital Animal Facility, Providence, RI), 8-12 weeks old, are maintained in an aseptic environment. Sixty-three animals are selected for the study such that body weights range from 25-30 grams the day prior to dosing. Animals are maintained in a quarantined room and handled under aseptic conditions. Food and water are supplied *ad libitum* throughout the experiment.

Appropriate tumor cells are obtained from the American Type Culture Collection (Rockville, MD) and are grown in modified Eagle's medium supplemented with 10% fetal calf serum. Five days prior to injection of the test material, mice receive a subcutaneous injection of tumor cells (approximately 2 x 10^6 cells/mouse) in the right rear flank.

Calipers are used to measure the dimensions of each tumor. Measurements (mm) of maximum and minimum width are performed prior to injection of the test material and at bi-weekly intervals for 61 days. Tumor volumes (mm^3) are computed using the formula Volume=[(minimum measurement)^2(maximum measurement)]/2. The results indicate that the HBEGF-containing conjugates substantially inhibit tumor cell proliferation *in vivo*.

EXAMPLE 11

PURIFICATION OF FPHS5 AND FPSH2

Purification steps 1 and 2 are performed with crude material being incubated on ice and all other steps are performed at room temperature using FPLC units/Biopilot (Pharmacia) equipped with P6000 pumps. Fractions/pools of the recombinant mitotoxins from the various chromatography steps are analyzed by SDS-PAGE.

Step 1: Preparation of cell extract. Cell paste (900 - 1200 g wet weight) is suspended in 3-4 volumes of ice-cold cell lysis buffer. (10mM sodium citrate, pH 6.0, containing 1 M urea, 5 mM EDTA, 5 mM EGTA and 50 mM NaCl) and passaged

3 times through a microfluidizer (Microfluidics Corp., Newton, MA, U.S.A.) at 18,000 lb/in^2. The resultant mixture was diluted to 10 volumes with lysis buffer.

Step 2: Expanded Bed Adsorption Chromatography (EBAC). Crude cell extract is loaded onto an expanded bed (300 ml of resin in a 5 x 100 cm column) of Streamline SP cation - exchange resin which is previously equilibrated with lysis buffer at 70 ml/min upwards flow. After loading, the resin is washed with the same buffer until the resin appears clear. The plunger is then slowly (1 - 2 cm/min) lowered. When the plunger nears the expanded bed and the A_{280} decreases to zero, the flow is stopped and the resin is allowed to settle. Once the plunger is 0.5 cm from the packed bed, proteins are eluted using 2 buffers containing increasing NaCl concentrations. The column is first washed with buffer A (10 mM sodium phosphate, pH 8.0, containing 5 mM EDTA, 5 mM EGTA and 0.25 M NaCl). After A_{280} reaches zero, buffer B (buffer A with 0.8 M NaCl) is applied. This eluate contains the conjugate which is subsequently diluted 1:3 (v/v) with buffer C (buffer A containing no NaCl) before being subjected to anion-exchange chromatography.

Step 3: Q-Sepharose anion-exchange chromatography and SP-Sepharose cation-exchange chromatography. Q-Sepharose removes contaminating *E. coli* proteins and DNA as well as endotoxin. The diluted HB-EGF-SAP pool from the previous step is loaded onto a column (2.6 x 7 cm) of Q-Sepharose FF directly connected to a column (2.6 x 13) of SP-sepharose HP. Both columns are previously equilibrated in tandem with buffer A containing 0.2 M NaCl. As the pI of the conjugate is above 9.5, it does not bind to the Q-Sepharose resin, but directly flows through and binds to S-Sepharose resin. When the A_{280} reaches zero, the anion-exchange column is disconnected from the cation-exchange column. Proteins bound to the S-Sepharose column are eluted with a gradient (10 column volumes) of 0.25 to 1 M NaCl in buffer A.

Step 4: Hydrophobic Interactions Chromatography (HIC). Fractions containing the conjugate are pooled, and solid ammonium sulfate is added (on ice and stirring) over a period of 15-20 min to a final concentration of 2 M. The pool is passed through a 0.8 μ filter and loaded onto a Phenyl-Sepharose HP column (2.6 x 10 cm) previously equilibrated with buffer D (10 mM sodium citrate, pH 6.0, containing 1 mM EGTA, 1 mM EGTA, 1 mM EDTA and 2 M $(NH_4)_2SO_4$). When the A_{280} reaches zero bound proteins are eluted using a gradient (10 column volumes) of 2 to 0 M $(NH_4)_2SO_4$ in the above buffer. Proteins in the various fractions are visualized by both SDS-PAGE

and Western blotting. There appears to be distinct pools of conjugates (3 in the case of FPHS5 and at least 2 in the case of FPSH2). In the case of FPHS5, two pools (A and B), are made. Monothioglycerol (MTG, 10 mM final concentration) is added to both pools. Pool A is dialyzed against formulation buffer (10 mM sodium citrate, pH 6.0, containing 0.1 mM EDTA and 0.14 M NaCl) and subjected to size-exclusion chromatography (Step 5). Pool B is dialyzed against buffer E (10 mM sodium phosphate buffer, pH 8.5, containing 5 mM EDTA and 5 mM EGTA) and subjected to another cation-exchange fractionation (Step 4).

Step 4B: S-Source cation-exchange chromatography. Pool B is applied to an S-Source column (2.6 x 5.7) previously equilibrated with buffer E and bound proteins are eluted with a gradient (10 column volumes) of 0 to 1 M NaCl in buffer E. Two pools (C and D) are made on the basis of SDS-PAGE analysis. Pool C contained an *E. coli* a major contaminant (~ 25-27 KDa), which is difficult to remove. Pool D is then subjected to size-exclusion chromatography (Step 5).

Step 5: Size exclusion chromatography. Both pools are passed separately through a suitably sized (i.e., the sample load volume is 10-15% of the total column volume) column containing S100 resin previously equilibrated with formulation buffer. From 1 kg. of wet weight paste, approximately 50 mg of purified FPHS5 (Pool A) and 10 mg of Pool D are recovered. At least 3 isoforms are apparent. Various analytical methods reveal the conjugates to be over 95% pure.

EXAMPLE 12

PURIFICATION OF FPHS2 AND FPHSH1

These fusion proteins were essentially purified in the same manner as in Example 1, except that a heparin sepharose FF affinity column was used prior to step 5.

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Barbara Sosnowski
Lois Chandler
L. L. Houston
Michael Nova
John R. McDonald

(ii) TITLE OF INVENTION: CONJUGATES OF HEPARIN-BINDING EPIDERMAL GROWTH FACTOR-LIKE GROWTH FACTOR AND TARGETED AGENTS

(iii) NUMBER OF SEQUENCES: 87

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Seed & Berry
(B) STREET: 6300 Columbia Center
(C) CITY: Seattle
(D) STATE: Washington
(E) COUNTRY: USA
(F) ZIP: 98104-7092

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/297,961
(B) FILING DATE: 29-AUG-1994
(C) CLASSIFICATION:

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/213,446
(B) FILING DATE: 15-MAR-1994
(C) CLASSIFICATION:

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/213,447
(B) FILING DATE: 15-MAR-1994
(C) CLASSIFICATION:

(2) INFORMATION FOR SEQ ID NO 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 627 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..627
(D) OTHER INFORMATION: /note "human HBEGF precursor"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
ATG AAG CTG CTG CCG TCG GTG GTG CTG AAG CTC TTT CTG GCT GCA GTT        48
Met Lys Leu Leu Pro Ser Val Val Leu Lys Leu Phe Leu Ala Ala Val
 1               5                  10                  15

CTC TCG GCA CTG GTG ACT GGC GAG AGC CTG GAG CGG CTT CGG AGA GGG        96
Leu Ser Ala Leu Val Thr Gly Glu Ser Leu Glu Arg Leu Arg Arg Gly
            20                  25                  30

CTA GCT GCT GGA ACC AGC AAC CCG GAC CCT CCC ACT GTA TCC ACG GAC       144
Leu Ala Ala Gly Thr Ser Asn Pro Asp Pro Pro Thr Val Ser Thr Asp
        35                  40                  45

CAG CTG CTA CCC CTA GGA GGC GGC CGG GAC CGG AAA GTC CGT GAC TTG       192
Gln Leu Leu Pro Leu Gly Gly Gly Arg Asp Arg Lys Val Arg Asp Leu
    50                  55                  60

CAA GAG GCA GAT CTG GAC CTT TTG AGA GTC ACT TTA TCC TCC AAG CCA       240
Gln Glu Ala Asp Leu Asp Leu Leu Arg Val Thr Leu Ser Ser Lys Pro
65                  70                  75                  80

CAA GCA CTG GCC ACA CCA AAC AAG GAG GAG CAC GGG AAA AGA AAG AAG       288
Gln Ala Leu Ala Thr Pro Asn Lys Glu Glu His Gly Lys Arg Lys Lys
                85                  90                  95

AAA GGC AAG GGG CTA GGG AAG AAG AGG GAC CCA TGT CTT CGG AAA TAC       336
Lys Gly Lys Gly Leu Gly Lys Lys Arg Asp Pro Cys Leu Arg Lys Tyr
            100                 105                 110

AAG GAC TTC TGC ATC CAT GGA GAA TGC AAA TAT GTG AAG GAG CTC CGG       384
Lys Asp Phe Cys Ile His Gly Glu Cys Lys Tyr Val Lys Glu Leu Arg
        115                 120                 125

GCT CCC TCC TGC ATC TGC CAC CCG GGT TAC CAT GGA GAG AGG TGT CAT       432
Ala Pro Ser Cys Ile Cys His Pro Gly Tyr His Gly Glu Arg Cys His
    130                 135                 140

GGG CTG AGC CTC CCA GTG GAA AAT CGC TTA TAT ACC TAT GAC CAC ACA       480
Gly Leu Ser Leu Pro Val Glu Asn Arg Leu Tyr Thr Tyr Asp His Thr
145                 150                 155                 160

ACC ATC CTG GCC GTG GTG GCT GTG GTG CTG TCA TCT GTC TGT CTG CTG       528
```

```
Thr Ile Leu Ala Val Val Ala Val Val Leu Ser Ser Val Cys Leu Leu
                165                 170                 175

GTC ATC GTG GGG CTT CTC ATG TTT AGG TAC CAT AGG AGA GGA GGT TAT      576
Val Ile Val Gly Leu Leu Met Phe Arg Tyr His Arg Arg Gly Gly Tyr
            180                 185                 190

GAT GTG GAA AAT GAA GAG AAA GTG AAG TTG GGC ATG ACT AAT TCC CAC      624
Asp Val Glu Asn Glu Glu Lys Val Lys Leu Gly Met Thr Asn Ser His
        195                 200                 205

TGA                                                                  627
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 208 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..627
(D) OTHER INFORMATION: /note "human HBEGF precursor"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Lys Leu Leu Pro Ser Val Val Leu Lys Leu Phe Leu Ala Ala Val
  1               5                  10                  15

Leu Ser Ala Leu Val Thr Gly Glu Ser Leu Glu Arg Leu Arg Arg Gly
            20                  25                  30

Leu Ala Ala Gly Thr Ser Asn Pro Asp Pro Pro Thr Val Ser Thr Asp
        35                  40                  45

Gln Leu Leu Pro Leu Gly Gly Gly Arg Asp Arg Lys Val Arg Asp Leu
    50                  55                  60

Gln Glu Ala Asp Leu Asp Leu Leu Arg Val Thr Leu Ser Ser Lys Pro
65                  70                  75                  80

Gln Ala Leu Ala Thr Pro Asn Lys Glu Glu His Gly Lys Arg Lys Lys
                85                  90                  95

Lys Gly Lys Gly Leu Gly Lys Lys Arg Asp Pro Cys Leu Arg Lys Tyr
            100                 105                 110

Lys Asp Phe Cys Ile His Gly Glu Cys Lys Tyr Val Lys Glu Leu Arg
        115                 120                 125
```

```
Ala Pro Ser Cys Ile Cys His Pro Gly Tyr His Gly Glu Arg Cys His
    130                 135                 140

Gly Leu Ser Leu Pro Val Glu Asn Arg Leu Tyr Thr Tyr Asp His Thr
145                 150                 155                 160

Thr Ile Leu Ala Val Val Ala Val Val Leu Ser Ser Val Cys Leu Leu
                165                 170                 175

Val Ile Val Gly Leu Leu Met Phe Arg Tyr His Arg Arg Gly Gly Tyr
            180                 185                 190

Asp Val Glu Asn Glu Glu Lys Val Lys Leu Gly Met Thr Asn Ser His
        195                 200                 205
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 77 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(ix) FEATURE:
(D) OTHER INFORMATION: /note "human mature HBEGF"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Arg Val Thr Leu Ser Ser Lys Pro Gln Ala Leu Ala Thr Pro Asn Lys
  1               5                  10                  15

Glu Glu His Gly Lys Arg Lys Lys Lys Gly Lys Gly Leu Gly Lys Lys
             20                  25                  30

Arg Asp Pro Cys Leu Arg Lys Tyr Lys Asp Phe Cys Ile His Gly Glu
         35                  40                  45

Cys Lys Tyr Val Lys Glu Leu Arg Ala Pro Ser Cys Ile Cys His Pro
     50                  55                  60

Gly Tyr His Gly Glu Arg Cys His Gly Leu Ser Leu Pro
 65                  70                  75
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 208 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(ix) FEATURE:
(D) OTHER INFORMATION: /note "monkey HBEGF precursor"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Lys Leu Leu Pro Ser Val Val Leu Lys Leu Leu Leu Ala Ala Val
1               5                  10                  15

Leu Ser Ala Leu Val Thr Gly Glu Ser Leu Glu Gln Leu Arg Arg Gly
            20                  25                  30

Leu Ala Ala Gly Thr Ser Asn Pro Asp Pro Ser Thr Gly Ser Thr Asp
        35                  40                  45

Gln Leu Leu Arg Leu Gly Gly Gly Arg Asp Arg Lys Val Arg Asp Leu
    50                  55                  60

Gln Glu Ala Asp Leu Asp Leu Leu Arg Val Thr Leu Ser Ser Lys Pro
65                  70                  75                  80

Gln Ala Leu Ala Thr Pro Ser Lys Glu Glu His Gly Lys Arg Lys Lys
                85                  90                  95

Lys Gly Lys Gly Leu Gly Lys Lys Arg Asp Pro Cys Leu Arg Lys Tyr
            100                 105                 110

Lys Asp Phe Cys Ile His Gly Glu Cys Lys Tyr Val Lys Glu Leu Arg
        115                 120                 125

Ala Pro Ser Cys Ile Cys His Pro Gly Tyr His Gly Glu Arg Cys His
    130                 135                 140

Gly Leu Ser Leu Pro Val Glu Asn Arg Leu Tyr Thr Tyr Asp His Thr
145                 150                 155                 160

Thr Ile Leu Ala Val Val Ala Val Val Leu Ser Ser Val Cys Leu Leu
                165                 170                 175

Val Ile Val Gly Leu Leu Met Phe Arg Tyr His Arg Arg Gly Gly Tyr
            180                 185                 190

Asp Val Glu Asn Glu Glu Lys Val Lys Leu Gly Met Thr Asn Ser His
        195                 200                 205
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 208 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(ix) FEATURE:
(D) OTHER INFORMATION: /note "rat HBEGF precursor"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Met Lys Leu Leu Pro Ser Val Val Leu Lys Leu Phe Leu Ala Ala Val
1               5                  10                  15

Leu Ser Ala Leu Val Thr Gly Glu Ser Leu Glu Arg Leu Arg Arg Gly
            20                  25                  30

Leu Ala Ala Ala Thr Ser Asn Pro Asp Pro Pro Thr Gly Thr Thr Asn
        35                  40                  45

Gln Leu Leu Pro Thr Gly Ala Asp Arg Ala Gln Glu Val Gln Asp Leu
    50                  55                  60

Glu Gly Thr Asp Leu Asp Leu Phe Lys Val Ala Phe Ser Ser Lys Pro
65                  70                  75                  80

Gln Ala Leu Ala Thr Pro Gly Lys Glu Lys Asn Gly Lys Lys Lys Arg
                85                  90                  95

Lys Gly Lys Gly Leu Gly Lys Lys Arg Asp Pro Cys Leu Lys Lys Tyr
            100                 105                 110

Lys Asp Tyr Cys Ile His Gly Glu Cys Arg Tyr Leu Lys Glu Leu Arg
        115                 120                 125

Ile Pro Ser Cys His Cys Leu Pro Gly Tyr His Gly Gln Arg Cys His
    130                 135                 140

Gly Leu Thr Leu Pro Val Glu Asn Pro Leu Tyr Thr Tyr Asp His Thr
145                 150                 155                 160

Thr Val Leu Ala Val Val Ala Val Val Leu Ser Ser Val Cys Leu Leu
                165                 170                 175

Val Ile Val Gly Leu Leu Met Phe Arg Tyr His Arg Arg Gly Gly Tyr
            180                 185                 190

Asp Leu Glu Ser Glu Glu Lys Val Lys Leu Gly Met Ala Ser Ser His
        195                 200                 205
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1002 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1002

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..627
(D) OTHER INFORMATION: /note "HBEGF Val Met Saporin"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
ATG AGA GTC ACT TTA TCC TCC AAG CCA CAA GCA CTG GCC ACA CCA AAC        48
Met Arg Val Thr Leu Ser Ser Lys Pro Gln Ala Leu Ala Thr Pro Asn
  1               5                  10                  15

AAG GAG GAG CAC GGG AAA AGA AAG AAG AAA GGC AAG GGG CTA GGG AAG        96
Lys Glu Glu His Gly Lys Arg Lys Lys Lys Gly Lys Gly Leu Gly Lys
            20                  25                  30

AAG AGG GAC CCA TGT CTT CGG AAA TAC AAG GAC TTC TGC ATC CAT GGA       144
Lys Arg Asp Pro Cys Leu Arg Lys Tyr Lys Asp Phe Cys Ile His Gly
        35                  40                  45

GAA TGC AAA TAT GTG AAG GAG CTC CGG GCT CCC TCC TGC ATC TGC CAC       192
Glu Cys Lys Tyr Val Lys Glu Leu Arg Ala Pro Ser Cys Ile Cys His
    50                  55                  60

CCG GGT TAC CAT GGA GAG AGG TGT CAT GGG CTG AGC CTC CCA GTC ATG       240
Pro Gly Tyr His Gly Glu Arg Cys His Gly Leu Ser Leu Pro Val Met
65                  70                  75                  80

GTC ACA TCA ATC ACA TTA GAT CTA GTA AAT CCG ACC GCG GGT CAA TAC       288
Val Thr Ser Ile Thr Leu Asp Leu Val Asn Pro Thr Ala Gly Gln Tyr
                85                  90                  95

TCA TCT TTT GTG GAT AAA ATC CGA AAC AAC GTA AAG GAT CCA AAC CTG       336
Ser Ser Phe Val Asp Lys Ile Arg Asn Asn Val Lys Asp Pro Asn Leu
            100                 105                 110

AAA TAC GGT GGT ACC GAC ATA GCC GTG ATA GGC CCA CCT TCT AAA GAA       384
Lys Tyr Gly Gly Thr Asp Ile Ala Val Ile Gly Pro Pro Ser Lys Glu
        115                 120                 125

AAA TTC CTT AGA ATT AAT TTC CAA AGT TCC CGA GGA ACG GTC TCA CTT       432
Lys Phe Leu Arg Ile Asn Phe Gln Ser Ser Arg Gly Thr Val Ser Leu
    130                 135                 140

GGC CTA AAA CGC GAT AAC TTG TAT GTG GTC GCG TAT CTT GCA ATG GAT       480
Gly Leu Lys Arg Asp Asn Leu Tyr Val Val Ala Tyr Leu Ala Met Asp
145                 150                 155                 160

AAC ACG AAT GTT AAT CGG GCA TAT TAC TTC AAA TCA GAA ATT ACT TCC       528
Asn Thr Asn Val Asn Arg Ala Tyr Tyr Phe Lys Ser Glu Ile Thr Ser
                165                 170                 175
```

```
GCC GAG TTA ACC GCC CTT TTC CCA GAG GCC ACA ACT GCA AAT CAG AAA        576
Ala Glu Leu Thr Ala Leu Phe Pro Glu Ala Thr Thr Ala Asn Gln Lys
            180                 185                 190

GCT TTA GAA TAC ACA GAA GAT TAT CAG TCG ATC GAA AAG AAT GCC CAG        624
Ala Leu Glu Tyr Thr Glu Asp Tyr Gln Ser Ile Glu Lys Asn Ala Gln
        195                 200                 205

ATA ACA CAG GGA GAT AAA AGT AGA AAA GAA CTC GGG TTG GGG ATC GAC        672
Ile Thr Gln Gly Asp Lys Ser Arg Lys Glu Leu Gly Leu Gly Ile Asp
    210                 215                 220

TTA CTT TTG ACG TTC ATG GAA GCA GTG AAC AAG AAG GCA CGT GTG GTT        720
Leu Leu Leu Thr Phe Met Glu Ala Val Asn Lys Lys Ala Arg Val Val
225                 230                 235                 240

AAA AAC GAA GCT AGG TTT CTG CTT ATC GCT ATT CAA ATG ACA GCT GAG        768
Lys Asn Glu Ala Arg Phe Leu Leu Ile Ala Ile Gln Met Thr Ala Glu
                245                 250                 255

GTA GCA CGA TTT AGG TAC ATT CAA AAC TTG GTA ACT AAG AAC TTC CCC        816
Val Ala Arg Phe Arg Tyr Ile Gln Asn Leu Val Thr Lys Asn Phe Pro
            260                 265                 270

AAC AAG TTC GAC TCG GAT AAC AAG GTG ATT CAA TTT GAA GTC AGC TGG        864
Asn Lys Phe Asp Ser Asp Asn Lys Val Ile Gln Phe Glu Val Ser Trp
        275                 280                 285

CGT AAG ATT TCT ACG GCA ATA TAC GGG GAT GCC AAA AAC GGC GTG TTT        912
Arg Lys Ile Ser Thr Ala Ile Tyr Gly Asp Ala Lys Asn Gly Val Phe
    290                 295                 300

AAT AAA GAT TAT GAT TTC GGG TTT GGA AAA GTG AGG CAG GTG AAG GAC        960
Asn Lys Asp Tyr Asp Phe Gly Phe Gly Lys Val Arg Gln Val Lys Asp
305                 310                 315                 320

TTG CAA ATG GGA CTC CTT ATG TAT TTG GGC AAA CCA AAG TAG               1002
Leu Gln Met Gly Leu Leu Met Tyr Leu Gly Lys Pro Lys
                325                 330
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
Ile Arg Val Arg Arg
  1               5
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 804 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..804

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..804
(D) OTHER INFORMATION: /note= "Nucleotide sequence corresponding to the clone M13 mp18-G4 in Example I.B.2."

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 46..804
(D) OTHER INFORMATION: /product= ""Saporin""

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
GCA TGG ATC CTG CTT CAA TTT TCA GCT TGG ACA ACA ACT GAT GCG GTC        48
Ala Trp Ile Leu Leu Gln Phe Ser Ala Trp Thr Thr Thr Asp Ala Val
-15                 -10                 -5                  1

ACA TCA ATC ACA TTA GAT CTA GTA AAT CCG ACC GCG GGT CAA TAC TCA        96
Thr Ser Ile Thr Leu Asp Leu Val Asn Pro Thr Ala Gly Gln Tyr Ser
             5                  10                  15

TCT TTT GTG GAT AAA ATC CGA AAC AAT GTA AAG GAT CCA AAC CTG AAA       144
Ser Phe Val Asp Lys Ile Arg Asn Asn Val Lys Asp Pro Asn Leu Lys
         20                  25                  30

TAC GGT GGT ACC GAC ATA GCC GTG ATA GGC CCA CCT TCT AAA GAA AAA       192
Tyr Gly Gly Thr Asp Ile Ala Val Ile Gly Pro Pro Ser Lys Glu Lys
     35                  40                  45

TTC CTT AGA ATT AAT TTC CAA AGT TCC CGA GGA ACG GTC TCA CTT GGC       240
Phe Leu Arg Ile Asn Phe Gln Ser Ser Arg Gly Thr Val Ser Leu Gly
 50                  55                  60                  65

CTA AAA CGC GAT AAC TTG TAT GTG GTC GCG TAT CTT GCA ATG GAT AAC       288
Leu Lys Arg Asp Asn Leu Tyr Val Val Ala Tyr Leu Ala Met Asp Asn
                 70                  75                  80

ACG AAT GTT AAT CGG GCA TAT TAC TTC AAA TCA GAA ATT ACT TCC GCC       336
Thr Asn Val Asn Arg Ala Tyr Tyr Phe Lys Ser Glu Ile Thr Ser Ala
             85                  90                  95
```

```
GAG TTA ACC GCC CTT TTC CCA GAG GCC ACA ACT GCA AAT CAG AAA GCT        384
Glu Leu Thr Ala Leu Phe Pro Glu Ala Thr Thr Ala Asn Gln Lys Ala
        100                 105                 110

TTA GAA TAC ACA GAA GAT TAT CAG TCG ATC GAA AAG AAT GCC CAG ATA        432
Leu Glu Tyr Thr Glu Asp Tyr Gln Ser Ile Glu Lys Asn Ala Gln Ile
    115                 120                 125

ACA CAG GGA GAT AAA AGT AGA AAA GAA CTC GGG TTG GGG ATC GAC TTA        480
Thr Gln Gly Asp Lys Ser Arg Lys Glu Leu Gly Leu Gly Ile Asp Leu
130                 135                 140                 145

CTT TTG ACG TTC ATG GAA GCA GTG AAC AAG AAG GCA CGT GTG GTT AAA        528
Leu Leu Thr Phe Met Glu Ala Val Asn Lys Lys Ala Arg Val Val Lys
                150                 155                 160

AAC GAA GCT AGG TTT CTG CTT ATC GCT ATT CAA ATG ACA GCT GAG GTA        576
Asn Glu Ala Arg Phe Leu Leu Ile Ala Ile Gln Met Thr Ala Glu Val
            165                 170                 175

GCA CGA TTT AGG TAC ATT CAA AAC TTG GTA ACT AAG AAC TTC CCC AAC        624
Ala Arg Phe Arg Tyr Ile Gln Asn Leu Val Thr Lys Asn Phe Pro Asn
        180                 185                 190

AAG TTC GAC TCG GAT AAC AAG GTG ATT CAA TTT GAA GTC AGC TGG CGT        672
Lys Phe Asp Ser Asp Asn Lys Val Ile Gln Phe Glu Val Ser Trp Arg
    195                 200                 205

AAG ATT TCT ACG GCA ATA TAC GGG GAT GCC AAA AAC GGC GTG TTT AAT        720
Lys Ile Ser Thr Ala Ile Tyr Gly Asp Ala Lys Asn Gly Val Phe Asn
210                 215                 220                 225

AAA GAT TAT GAT TTC GGG TTT GGA AAA GTG AGG CAG GTG AAG GAC TTG        768
Lys Asp Tyr Asp Phe Gly Phe Gly Lys Val Arg Gln Val Lys Asp Leu
                230                 235                 240

CAA ATG GGA CTC CTT ATG TAT TTG GGC AAA CCA AAG                        804
Gln Met Gly Leu Leu Met Tyr Leu Gly Lys Pro Lys
            245                 250
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 804 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..804

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..804
(D) OTHER INFORMATION: /note= "Nucleotide sequence corresponding to the clone M13 mp18-G1 in Example I.B.2."

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 46..804
(D) OTHER INFORMATION: /product= "Saporin"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
GCA TGG ATC CTG CTT CAA TTT TCA GCT TGG ACA ACA ACT GAT GCG GTC        48
Ala Trp Ile Leu Leu Gln Phe Ser Ala Trp Thr Thr Thr Asp Ala Val
-15                 -10                 -5                  1

ACA TCA ATC ACA TTA GAT CTA GTA AAT CCG ACC GCG GGT CAA TAC TCA        96
Thr Ser Ile Thr Leu Asp Leu Val Asn Pro Thr Ala Gly Gln Tyr Ser
             5                  10                  15

TCT TTT GTG GAT AAA ATC CGA AAC AAC GTA AAG GAT CCA AAC CTG AAA       144
Ser Phe Val Asp Lys Ile Arg Asn Asn Val Lys Asp Pro Asn Leu Lys
         20                  25                  30

TAC GGT GGT ACC GAC ATA GCC GTG ATA GGC CCA CCT TCT AAA GAA AAA       192
Tyr Gly Gly Thr Asp Ile Ala Val Ile Gly Pro Pro Ser Lys Glu Lys
     35                  40                  45


TTC CTT AGA ATT AAT TTC CAA AGT TCC CGA GGA ACG GTC TCA CTT GGC       240
Phe Leu Arg Ile Asn Phe Gln Ser Ser Arg Gly Thr Val Ser Leu Gly
 50                  55                  60                  65

CTA AAA CGC GAT AAC TTG TAT GTG GTC GCG TAT CTT GCA ATG GAT AAC       288
Leu Lys Arg Asp Asn Leu Tyr Val Val Ala Tyr Leu Ala Met Asp Asn
                 70                  75                  80

ACG AAT GTT AAT CGG GCA TAT TAC TTC AGA TCA GAA ATT ACT TCC GCC       336
Thr Asn Val Asn Arg Ala Tyr Tyr Phe Arg Ser Glu Ile Thr Ser Ala
             85                  90                  95

GAG TTA ACC GCC CTT TTC CCA GAG GCC ACA ACT GCA AAT CAG AAA GCT       384
Glu Leu Thr Ala Leu Phe Pro Glu Ala Thr Thr Ala Asn Gln Lys Ala
        100                 105                 110

TTA GAA TAC ACA GAA GAT TAT CAG TCG ATC GAA AAG AAT GCC CAG ATA       432
Leu Glu Tyr Thr Glu Asp Tyr Gln Ser Ile Glu Lys Asn Ala Gln Ile
    115                 120                 125

ACA CAG GGA GAT AAA TCA AGA AAA GAA CTC GGG TTG GGG ATC GAC TTA       480
Thr Gln Gly Asp Lys Ser Arg Lys Glu Leu Gly Leu Gly Ile Asp Leu
130                 135                 140                 145
```

```
CTT TTG ACG TCC ATG GAA GCA GTG AAC AAG AAG GCA CGT GTG GTT AAA        528
Leu Leu Thr Ser Met Glu Ala Val Asn Lys Lys Ala Arg Val Val Lys
                150                 155                 160

AAC GAA GCT AGG TTT CTG CTT ATC GCT ATT CAA ATG ACA GCT GAG GTA        576
Asn Glu Ala Arg Phe Leu Leu Ile Ala Ile Gln Met Thr Ala Glu Val
            165                 170                 175

GCA CGA TTT CGG TAC ATT CAA AAC TTG GTA ACT AAG AAC TTC CCC AAC        624
Ala Arg Phe Arg Tyr Ile Gln Asn Leu Val Thr Lys Asn Phe Pro Asn
        180                 185                 190

AAG TTC GAC TCG GAT AAC AAG GTG ATT CAA TTT GAA GTC AGC TGG CGT        672
Lys Phe Asp Ser Asp Asn Lys Val Ile Gln Phe Glu Val Ser Trp Arg
    195                 200                 205

AAG ATT TCT ACG GCA ATA TAC GGA GAT GCC AAA AAC GGC GTG TTT AAT        720
Lys Ile Ser Thr Ala Ile Tyr Gly Asp Ala Lys Asn Gly Val Phe Asn
210                 215                 220                 225

AAA GAT TAT GAT TTC GGG TTT GGA AAA GTG AGG CAG GTG AAG GAC TTG        768
Lys Asp Tyr Asp Phe Gly Phe Gly Lys Val Arg Gln Val Lys Asp Leu
                230                 235                 240

CAA ATG GGA CTC CTT ATG TAT TTG GGC AAA CCA AAG                        804
Gln Met Gly Leu Leu Met Tyr Leu Gly Lys Pro Lys
            245                 250
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 804 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..804

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..804
(D) OTHER INFORMATION: /note= "Nucleotide sequence corresponding to the clone M13 mp18-G2 in Example I.B.2."

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 46..804
(D) OTHER INFORMATION: /product= "Saporin"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
GCA TGG ATC CTG CTT CAA TTT TCA GCT TGG ACA ACA ACT GAT GCG GTC        48
Ala Trp Ile Leu Leu Gln Phe Ser Ala Trp Thr Thr Thr Asp Ala Val
-15                 -10                 -5                  1

ACA TCA ATC ACA TTA GAT CTA GTA AAT CCG ACT GCG GGT CAA TAC TCA        96
Thr Ser Ile Thr Leu Asp Leu Val Asn Pro Thr Ala Gly Gln Tyr Ser
              5                  10                  15

TCT TTT GTG GAT AAA ATC CGA AAC AAC GTA AAG GAT CCA AAC CTG AAA       144
Ser Phe Val Asp Lys Ile Arg Asn Asn Val Lys Asp Pro Asn Leu Lys
          20                  25                  30

TAC GGT GGT ACC GAC ATA GCC GTG ATA GGC CCA CCT TCT AAA GAT AAA       192
Tyr Gly Gly Thr Asp Ile Ala Val Ile Gly Pro Pro Ser Lys Asp Lys
      35                  40                  45

TTC CTT AGA ATT AAT TTC CAA AGT TCC CGA GGA ACG GTC TCA CTT GGC       240
Phe Leu Arg Ile Asn Phe Gln Ser Ser Arg Gly Thr Val Ser Leu Gly
 50                  55                  60                  65

CTA AAA CGC GAT AAC TTG TAT GTG GTC GCG TAT CTT GCA ATG GAT AAC       288
Leu Lys Arg Asp Asn Leu Tyr Val Val Ala Tyr Leu Ala Met Asp Asn
                 70                  75                  80

ACG AAT GTT AAT CGG GCA TAT TAC TTC AAA TCA GAA ATT ACT TCC GCC       336
Thr Asn Val Asn Arg Ala Tyr Tyr Phe Lys Ser Glu Ile Thr Ser Ala
             85                  90                  95

GAG TTA ACC GCC CTT TTC CCA GAG GCC ACA ACT GCA AAT CAG AAA GCT       384
Glu Leu Thr Ala Leu Phe Pro Glu Ala Thr Thr Ala Asn Gln Lys Ala
         100                 105                 110

TTA GAA TAC ACA GAA GAT TAT CAG TCG ATC GAA AAG AAT GCC CAG ATA       432
Leu Glu Tyr Thr Glu Asp Tyr Gln Ser Ile Glu Lys Asn Ala Gln Ile
     115                 120                 125

ACA CAG GGA GAT AAA AGT AGA AAA GAA CTC GGG TTG GGG ATC GAC TTA       480
Thr Gln Gly Asp Lys Ser Arg Lys Glu Leu Gly Leu Gly Ile Asp Leu
130                 135                 140                 145

CTT TTG ACG TTC ATG GAA GCA GTG AAC AAG AAG GCA CGT GTG GTT AAA       528
Leu Leu Thr Phe Met Glu Ala Val Asn Lys Lys Ala Arg Val Val Lys
                150                 155                 160

AAC GAA GCT AGG TTT CTG CTT ATC GCT ATT CAA ATG ACA GCT GAG GTA       576
Asn Glu Ala Arg Phe Leu Leu Ile Ala Ile Gln Met Thr Ala Glu Val
            165                 170                 175

GCA CGA TTT AGG TAC ATT CAA AAC TTG GTA ACT AAG AAC TTC CCC AAC       624
Ala Arg Phe Arg Tyr Ile Gln Asn Leu Val Thr Lys Asn Phe Pro Asn
        180                 185                 190

AAG TTC GAC TCG GAT AAC AAG GTG ATT CAA TTT GAA GTC AGC TGG CGT       672
Lys Phe Asp Ser Asp Asn Lys Val Ile Gln Phe Glu Val Ser Trp Arg
    195                 200                 205
```

```
AAG ATT TCT ACG GCA ATA TAC GGG GAT GCC AAA AAC GGC GTG TTT AAT        720
Lys Ile Ser Thr Ala Ile Tyr Gly Asp Ala Lys Asn Gly Val Phe Asn
210                 215                 220                 225

AAA GAT TAT GAT TTC GGG TTT GGA AAA GTG AGG CAG GTG AAG GAC TTG        768
Lys Asp Tyr Asp Phe Gly Phe Gly Lys Val Arg Gln Val Lys Asp Leu
                230                 235                 240

CAA ATG GGA CTC CTT ATG TAT TTG GGC AAA CCA AAG                        804
Gln Met Gly Leu Leu Met Tyr Leu Gly Lys Pro Lys
            245                 250
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 804 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..804

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..804
(D) OTHER INFORMATION: /note= "Nucleotide sequence corresponding to the clone M13 mp18-G7 in Example I.B.2."

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 46..804
(D) OTHER INFORMATION: /product= "Saporin"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
GCA TGG ATC CTG CTT CAA TTT TCA GCT TGG ACA ACA ACT GAT GCG GTC         48
Ala Trp Ile Leu Leu Gln Phe Ser Ala Trp Thr Thr Thr Asp Ala Val
-15                 -10                 -5                  1

ACA TCA ATC ACA TTA GAT CTA GTA AAT CCG ACC GCG GGT CAA TAC TCA         96
Thr Ser Ile Thr Leu Asp Leu Val Asn Pro Thr Ala Gly Gln Tyr Ser
            5                  10                  15

TCT TTT GTG GAT AAA ATC CGA AAC AAC GTA AAG GAT CCA AAC CTG AAA        144
Ser Phe Val Asp Lys Ile Arg Asn Asn Val Lys Asp Pro Asn Leu Lys
        20                  25                  30

TAC GGT GGT ACC GAC ATA GCC GTG ATA GGC CCA CCT TCT AAA GAA AAA        192
```

```
Tyr Gly Gly Thr Asp Ile Ala Val Ile Gly Pro Pro Ser Lys Glu Lys
     35                  40                  45

TTC CTT AGA ATT AAT TTC CAA AGT TCC CGA GGA ACG GTC TCA CTT GGC       240
Phe Leu Arg Ile Asn Phe Gln Ser Ser Arg Gly Thr Val Ser Leu Gly
 50                  55                  60                  65

CTA AAA CGC GAT AAC TTG TAT GTG GTC GCG TAT CTT GCA ATG GAT AAC       288
Leu Lys Arg Asp Asn Leu Tyr Val Val Ala Tyr Leu Ala Met Asp Asn
                 70                  75                  80

ACG AAT GTT AAT CGG GCA TAT TAC TTC AGA TCA GAA ATT ACT TCC GCC       336
Thr Asn Val Asn Arg Ala Tyr Tyr Phe Arg Ser Glu Ile Thr Ser Ala
             85                  90                  95

GAG TTA ACC GCC CTT TTC CCA GAG GCC ACA ACT GCA AAT CAG AAA GCT       384
Glu Leu Thr Ala Leu Phe Pro Glu Ala Thr Thr Ala Asn Gln Lys Ala
        100                 105                 110

TTA GAA TAC ACA GAA GAT TAT CAG TCG ATC GAA AAG AAT GCC CAG ATA       432
Leu Glu Tyr Thr Glu Asp Tyr Gln Ser Ile Glu Lys Asn Ala Gln Ile
    115                 120                 125

ACA CAG GGA GAT AAA TCA AGA AAA GAA CTC GGG TTG GGG ATC GAC TTA       480
Thr Gln Gly Asp Lys Ser Arg Lys Glu Leu Gly Leu Gly Ile Asp Leu
130                 135                 140                 145

CTT TTG ACG TCC ATG GAA GCA GTG AAC AAG AAG GCA CGT GTG GTT AAA       528
Leu Leu Thr Ser Met Glu Ala Val Asn Lys Lys Ala Arg Val Val Lys
                150                 155                 160

AAC GAA GCT AGA TTC CTT CTT ATC GCT ATT CAG ATG ACG GCT GAG GCA       576
Asn Glu Ala Arg Phe Leu Leu Ile Ala Ile Gln Met Thr Ala Glu Ala
            165                 170                 175

GCA CGA TTT AGG TAC ATA CAA AAC TTG GTA ATC AAG AAC TTT CCC AAC       624
Ala Arg Phe Arg Tyr Ile Gln Asn Leu Val Ile Lys Asn Phe Pro Asn
        180                 185                 190

AAG TTC AAC TCG GAA AAC AAA GTG ATT CAG TTT GAG GTT AAC TGG AAA       672
Lys Phe Asn Ser Glu Asn Lys Val Ile Gln Phe Glu Val Asn Trp Lys
    195                 200                 205

AAA ATT TCT ACG GCA ATA TAC GGG GAT GCC AAA AAC GGC GTG TTT AAT       720
Lys Ile Ser Thr Ala Ile Tyr Gly Asp Ala Lys Asn Gly Val Phe Asn
210                 215                 220                 225

AAA GAT TAT GAT TTC GGG TTT GGA AAA GTG AGG CAG GTG AAG GAC TTG       768
Lys Asp Tyr Asp Phe Gly Phe Gly Lys Val Arg Gln Val Lys Asp Leu
                230                 235                 240

CAA ATG GGA CTC CTT ATG TAT TTG GGC AAA CCA AAG                       804
Gln Met Gly Leu Leu Met Tyr Leu Gly Lys Pro Lys
            245                 250
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 804 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..804

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..804
(D) OTHER INFORMATION: /note= "Nucleotide sequence corresponding to the clone M13 mp18-G9 in Example I.B.2."

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 46..804
(D) OTHER INFORMATION: /product= "Saporin"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
GCA TGG ATC CTG CTT CAA TTT TCA GCT TGG ACA ACA ACT GAT GCG GTC        48
Ala Trp Ile Leu Leu Gln Phe Ser Ala Trp Thr Thr Thr Asp Ala Val
-15                 -10                 -5                  1

ACA TCA ATC ACA TTA GAT CTA GTA AAT CCG ACC GCG GGT CAA TAC TCA        96
Thr Ser Ile Thr Leu Asp Leu Val Asn Pro Thr Ala Gly Gln Tyr Ser
             5                  10                  15

TCT TTT GTG GAT AAA ATC CGA AAC AAC GTA AAG GAT CCA AAC CTG AAA       144
Ser Phe Val Asp Lys Ile Arg Asn Asn Val Lys Asp Pro Asn Leu Lys
        20                  25                  30

TAC GGT GGT ACC GAC ATA GCC GTG ATA GGC CCA CCT TCT AAA GAA AAA       192
Tyr Gly Gly Thr Asp Ile Ala Val Ile Gly Pro Pro Ser Lys Glu Lys
    35                  40                  45

TTC CTT AGA ATT AAT TTC CAA AGT TCC CGA GGA ACG GTC TCA CTT GGC       240
Phe Leu Arg Ile Asn Phe Gln Ser Ser Arg Gly Thr Val Ser Leu Gly
50                  55                  60                  65

CTA AAA CGC GAT AAC TTG TAT GTG GTC GCG TAT CTT GCA ATG GAT AAC       288
Leu Lys Arg Asp Asn Leu Tyr Val Val Ala Tyr Leu Ala Met Asp Asn
                70                  75                  80

ACG AAT GTT AAT CGG GCA TAT TAC TTC AGA TCA GAA ATT ACT TCC GCC       336
Thr Asn Val Asn Arg Ala Tyr Tyr Phe Arg Ser Glu Ile Thr Ser Ala
            85                  90                  95
```

```
GAG TTA ACC GCC CTT TTC CCA GAG GCC ACA ACT GCA AAT CAG AAA GCT        384
Glu Leu Thr Ala Leu Phe Pro Glu Ala Thr Thr Ala Asn Gln Lys Ala
        100                 105                 110

TTA GAA TAC ACA GAA GAT TAT CAG TCG ATT GAA AAG AAT GCC CAG ATA        432
Leu Glu Tyr Thr Glu Asp Tyr Gln Ser Ile Glu Lys Asn Ala Gln Ile
    115                 120                 125

ACA CAA GGA GAT CAA AGT AGA AAA GAA CTC GGG TTG GGG ATT GAC TTA        480
Thr Gln Gly Asp Gln Ser Arg Lys Glu Leu Gly Leu Gly Ile Asp Leu
130                 135                 140                 145

CTT TCA ACG TCC ATG GAA GCA GTG AAC AAG AAG GCA CGT GTG GTT AAA        528
Leu Ser Thr Ser Met Glu Ala Val Asn Lys Lys Ala Arg Val Val Lys
                150                 155                 160

GAC GAA GCT AGA TTC CTT CTT ATC GCT ATT CAG ATG ACG GCT GAG GCA        576
Asp Glu Ala Arg Phe Leu Leu Ile Ala Ile Gln Met Thr Ala Glu Ala
            165                 170                 175

GCG CGA TTT AGG TAC ATA CAA AAC TTG GTA ATC AAG AAC TTT CCC AAC        624
Ala Arg Phe Arg Tyr Ile Gln Asn Leu Val Ile Lys Asn Phe Pro Asn
        180                 185                 190

AAG TTC AAC TCG GAA AAC AAA GTG ATT CAG TTT GAG GTT AAC TGG AAA        672
Lys Phe Asn Ser Glu Asn Lys Val Ile Gln Phe Glu Val Asn Trp Lys
    195                 200                 205

AAA ATT TCT ACG GCA ATA TAC GGG GAT GCC AAA AAC GGC GTG TTT AAT        720
Lys Ile Ser Thr Ala Ile Tyr Gly Asp Ala Lys Asn Gly Val Phe Asn
210                 215                 220                 225

AAA GAT TAT GAT TTC GGG TTT GGA AAA GTG AGG CAG GTG AAG GAC TTG        768
Lys Asp Tyr Asp Phe Gly Phe Gly Lys Val Arg Gln Val Lys Asp Leu
                230                 235                 240

CAA ATG GGA CTC CTT ATG TAT TTG GGC AAA CCA AAG                        804
Gln Met Gly Leu Leu Met Tyr Leu Gly Lys Pro Lys
            245                 250
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(D) OTHER INFORMATION: /product= "SO-4"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Val Ile Ile Tyr Glu Leu Asn Leu Gln Gly Thr Thr Lys Ala Gln Tyr
5 10 15

Ser Thr Ile Leu Lys Gln Leu Arg Asp Asp Ile Lys Asp Pro Asn Leu
20 25 30

Xaa Tyr Gly Xaa Xaa Asp Tyr Ser
35 40

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Ile Lys Arg Gln Arg Arg
1 5

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CATATGTGTG TCACATCAAT CACATTAGAT 30

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CAGGTTTGGA TCCTTTACGT T 21

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: misc_recomb
(B) LOCATION: 10..15
(D) OTHER INFORMATION: /standard_name= "Nco I restriction enzyme recognition site"

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 15..22
(D) OTHER INFORMATION: /product= "N-terminus of Saporin protein"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CAACAACTGC CATGGTCACA TC 22

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 59 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(D) OTHER INFORMATION: /product= trp promoter

AATTCCCCTG TTGACAATTA ATCATCGAAC TAGTTAACTA GTACGCAGCT TGGCTGCAG 59

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 59 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(D) OTHER INFORMATION/product= bacteriophage lambda CII ribosome binding site

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GTCGACCAAG CTTGGGCATA CATTCAATCA ATTGTTATCT AAGGAAATAC TTACATATG 59

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: misc_recomb
(B) LOCATION: 11..16
(D) OTHER INFORMATION: /standard_name= "Nco I restriction enzyme rccognition site."

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 1..10
(D) OTHER INFORMATION: /product= "Carboxy terminus of mature FGF protein"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GCT AAG AGC GCC ATG GAGA 19
Ala Lys Ser Ala Met
1 5

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..12
(D) OTHER INFORMATION: /product= "Carboxy terminus of wild type FGF"

(ix) FEATURE:
(A) NAME/KEY: misc_recomb
(B) LOCATION: 13..18
(D) OTHER INFORMATION: /standard_name= "Nco I restriction enzyme recognition site"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```
GCT AAG AGC TGACCATGGA GA                                              21
Ala Lys Ser
  1

(2) INFORMATION FOR SEQ ID NO:22:

     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 102 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: double
          (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (ix) FEATURE:
          (A) NAME/KEY: CDS
          (B) LOCATION: 1..96
          (D) OTHER INFORMATION: /product= "pFGFNcoI"
                /note= "Equals the plasmid pFC80 with native FGF
                stop codon removed."

    (ix) FEATURE:
          (A) NAME/KEY: misc_recomb
          (B) LOCATION: 29..34
           (D) OTHER INFORMATION: /standard_name= "Nco I restriction enzyme
recognition site"

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CTT TTT CTT CCA ATG TCT GCT AAG AGC GCC ATG GAG ATC CGG CTG AAT        48
Leu Phe Leu Pro Met Ser Ala Lys Ser Ala Met Glu Ile Arg Leu Asn
  1               5                  10                  15

GGT GCA GTT CTG TAC CGG TTT TCC TGT GCC GTC TTT CAG GAC TCC TGAAATCTT
102
Gly Ala Val Leu Tyr Arg Phe Ser Cys Ala Val Phe Gln Asp Ser
             20                  25                  30

(2) INFORMATION FOR SEQ ID NO:23:

     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 88 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

AAGGAGATATACC ATG GGC AGC AGC CAT CAT CAT CAT CAT CAC AGC AGC          43
              Met Gly Ser Ser His His His His His His Ser Ser
               1               5                  10

GGC CTG GTG CCG CGC GGC AGC CAT ATG CTC GAG GAT CCG                    88
```

Gly Leu Val Pro Arg Gly Ser His Met Leu Glu Asp Pro
15 20 25

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GGATCCGCCT CGTTTGACTA CTT 23

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CTGGACCATA TGAGAGTCAC TTTA 24

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GTATATCATG ACTGGGAGGC TCAGCCCATG ACA 33

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: misc_recomb
(B) LOCATION: 6..11
(D) OTHER INFORMATION: /standard_name= "EcoRI Restriction Site"

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 12..30
(D) OTHER INFORMATION: /function= "N-terminal extension"
/product= "Native saporin signal peptide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

CTGCAGAATT CGCATGGATC CTGCTTCAAT 30

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(ix) FEATURE:
(A) NAME/KEY: misc_recomb
(B) LOCATION: 6..11
(D) OTHER INFORMATION: /standard_name= "EcoRI Restriction Site"

(ix) FEATURE:
(A) NAME/KEY: terminator
(B) LOCATION: 23..25
(D) OTHER INFORMATION: /note= "Anti-sense stop codon"

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 26..30
(D) OTHER INFORMATION: /note= "Anti-sense to carboxyl
terminus of mature peptide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CTGCAGAATT CGCCTCGTTT GACTACTTTG 30

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 46 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

AGCCCGGAGC TCCTTCACAT ATTTGCATTC TCCGTGGATG CAGAAG 46

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GTGAAGGAGC TCCGGGCTCC CTCCTGCATC TGCCACCCGG GTTATCATGG AGAGAGG 57

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

ATATAGAATT CTGTCTTCTC AGAGGTA 27

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

ATATACCATG GCTGGGAGGC TCAGCCCATG ACA 33

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1002 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1002
(D) OTHER INFORMATION: /product= "Linker Amenable HBEGF-SAP"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33

```
ATG AGA GTC ACT TTA TCC TCC AAG CCA CAA GCA CTG GCC ACA CCA AAC        48
Met Arg Val Thr Leu Ser Ser Lys Pro Gln Ala Leu Ala Thr Pro Asn
 1               5                  10                  15

AAG GAG GAG CAC GGG AAA AGA AAG AAG AAA GGC AAG GGG CTA GGG AAG        96
Lys Glu Glu His Gly Lys Arg Lys Lys Lys Gly Lys Gly Leu Gly Lys
            20                  25                  30

AAG AGG GAC CCA TGT CTT CGG AAA TAC AAG GAC TTC TGC ATC CAC GGA       144
Lys Arg Asp Pro Cys Leu Arg Lys Tyr Lys Asp Phe Cys Ile His Gly
        35                  40                  45

GAA TGC AAA TAT GTG AAG GAG CTC CGG GCT CCC TCC TGC ATC TGC CAC       192
Glu Cys Lys Tyr Val Lys Glu Leu Arg Ala Pro Ser Cys Ile Cys His
    50                  55                  60

CCG GGT TAT CAT GGA GAG AGG TGT CAT GGG CTG AGC CTC CCA GCC ATG       240
Pro Gly Tyr His Gly Glu Arg Cys His Gly Leu Ser Leu Pro Ala Met
65                  70                  75                  80

GTC ACA TCA ATC ACA TTA GAT CTA GTA AAT CCG ACC GCG GGT CAA TAC       288
Val Thr Ser Ile Thr Leu Asp Leu Val Asn Pro Thr Ala Gly Gln Tyr
                85                  90                  95

TCA TCT TTT GTG GAT AAA ATC CGA AAC AAC GTA AAG GAT CCA AAC CTG       336
Ser Ser Phe Val Asp Lys Ile Arg Asn Asn Val Lys Asp Pro Asn Leu
            100                 105                 110

AAA TAC GGT GGT ACC GAC ATA GCC GTG ATA GGC CCA CCT TCT AAA GAA       384
Lys Tyr Gly Gly Thr Asp Ile Ala Val Ile Gly Pro Pro Ser Lys Glu
        115                 120                 125

AAA TTC CTT AGA ATT AAT TTC CAA AGT TCC CGA GGA ACG GTC TCA CTT       432
Lys Phe Leu Arg Ile Asn Phe Gln Ser Ser Arg Gly Thr Val Ser Leu
```

```
    130                 135                 140

GGC CTA AAA CGC GAT AAC TTG TAT GTG GTC GCG TAT CTT GCA ATG GAT      480
Gly Leu Lys Arg Asp Asn Leu Tyr Val Val Ala Tyr Leu Ala Met Asp
145                 150                 155                 160

AAC ACG AAT GTT AAT CGG GCA TAT TAC TTC AAA TCA GAA ATT ACT TCC      528
Asn Thr Asn Val Asn Arg Ala Tyr Tyr Phe Lys Ser Glu Ile Thr Ser
                165                 170                 175

GCC GAG TTA ACC GCC CTT TTC CCA GAG GCC ACA ACT GCA AAT CAG AAA      576
Ala Glu Leu Thr Ala Leu Phe Pro Glu Ala Thr Thr Ala Asn Gln Lys
            180                 185                 190

GCT TTA GAA TAC ACA GAA GAT TAT CAG TCG ATC GAA AAG AAT GCC CAG      624
Ala Leu Glu Tyr Thr Glu Asp Tyr Gln Ser Ile Glu Lys Asn Ala Gln
        195                 200                 205

ATA ACA CAG GGA GAT AAA AGT AGA AAA GAA CTC GGG TTG GGG ATC GAC      672
Ile Thr Gln Gly Asp Lys Ser Arg Lys Glu Leu Gly Leu Gly Ile Asp
    210                 215                 220

TTA CTT TTG ACG TTC ATG GAA GCA GTG AAC AAG AAG GCA CGT GTG GTT      720
Leu Leu Leu Thr Phe Met Glu Ala Val Asn Lys Lys Ala Arg Val Val
225                 230                 235                 240

AAA AAC GAA GCT AGG TTT CTG CTT ATC GCT ATT CAA ATG ACA GCT GAG      768
Lys Asn Glu Ala Arg Phe Leu Leu Ile Ala Ile Gln Met Thr Ala Glu
                245                 250                 255

GTA GCA CGA TTT AGG TAC ATT CAA AAC TTG GTA ACT AAG AAC TTC CCC      816
Val Ala Arg Phe Arg Tyr Ile Gln Asn Leu Val Thr Lys Asn Phe Pro
            260                 265                 270

AAC AAG TTC GAC TCG GAT AAC AAG GTG ATT CAA TTT GAA GTC AGC TGG      864
Asn Lys Phe Asp Ser Asp Asn Lys Val Ile Gln Phe Glu Val Ser Trp
        275                 280                 285

CGT AAG ATT TCT ACG GCA ATA TAC GGG GAT GCC AAA AAC GGC GTG TTT      912
Arg Lys Ile Ser Thr Ala Ile Tyr Gly Asp Ala Lys Asn Gly Val Phe
    290                 295                 300

AAT AAA GAT TAT GAT TTC GGG TTT GGA AAA GTG AGG CAG GTG AAG GAC      960
Asn Lys Asp Tyr Asp Phe Gly Phe Gly Lys Val Arg Gln Val Lys Asp
305                 310                 315                 320

TTG CAA ATG GGA CTC CTT ATG TAT TTG GGC AAA CCA AAG TAG             1002
Leu Gln Met Gly Leu Leu Met Tyr Leu Gly Lys Pro Lys
                325                 330
```

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1230 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1230

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 1..465
(D) OTHER INFORMATION: /product= "bFGF"

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 472..1230
(D) OTHER INFORMATION: /product= "Saporin"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```
ATG GCA GCA GGA TCA ATA ACA ACA TTA CCC GCC TTG CCC GAG GAT GGC        48
Met Ala Ala Gly Ser Ile Thr Thr Leu Pro Ala Leu Pro Glu Asp Gly
  1               5                  10                  15

GGC AGC GGC GCC TTC CCG CCC GGC CAC TTC AAG GAC CCC AAG CGG CTG        96
Gly Ser Gly Ala Phe Pro Pro Gly His Phe Lys Asp Pro Lys Arg Leu
             20                  25                  30

TAC TGC AAA AAC GGG GGC TTC TTC CTG CGC ATC CAC CCC GAC GGC CGA       144
Tyr Cys Lys Asn Gly Gly Phe Phe Leu Arg Ile His Pro Asp Gly Arg
         35                  40                  45

GTT GAC GGG GTC CGG GAG AAG AGC GAC CCT CAC ATC AAG CTT CAA CTT       192
Val Asp Gly Val Arg Glu Lys Ser Asp Pro His Ile Lys Leu Gln Leu
     50                  55                  60

CAA GCA GAA GAG AGA GGA GTT GTG TCT ATC AAA GGA GTG TGT GCT AAC       240
Gln Ala Glu Glu Arg Gly Val Val Ser Ile Lys Gly Val Cys Ala Asn
 65                  70                  75                  80

CGT TAC CTG GCT ATG AAG GAA GAT GGA AGA TTA CTG GCT TCT AAA TGT       288
Arg Tyr Leu Ala Met Lys Glu Asp Gly Arg Leu Leu Ala Ser Lys Cys
                 85                  90                  95

GTT ACG GAT GAG TGT TTC TTT TTT GAA CGA TTG GAA TCT AAT AAC TAC       336
Val Thr Asp Glu Cys Phe Phe Phe Glu Arg Leu Glu Ser Asn Asn Tyr
            100                 105                 110

AAT ACT TAC CGG TCA AGG AAA TAC ACC AGT TGG TAT GTG GCA TTG AAA       384
Asn Thr Tyr Arg Ser Arg Lys Tyr Thr Ser Trp Tyr Val Ala Leu Lys
        115                 120                 125

CGA ACT GGG CAG TAT AAA CTT GGA TCC AAA ACA GGA CCT GGG CAG AAA       432
Arg Thr Gly Gln Tyr Lys Leu Gly Ser Lys Thr Gly Pro Gly Gln Lys
```

```
    130                 135                 140

GCT ATA CTT TTT CTT CCA ATG TCT GCT AAG AGC GCC ATG GTC ACA TCA        480
Ala Ile Leu Phe Leu Pro Met Ser Ala Lys Ser Ala Met Val Thr Ser
145                 150                 155                 160

ATC ACA TTA GAT CTA GTA AAT CCG ACC GCG GGT CAA TAC TCA TCT TTT        528
Ile Thr Leu Asp Leu Val Asn Pro Thr Ala Gly Gln Tyr Ser Ser Phe
                165                 170                 175

GTG GAT AAA ATC CGA AAC AAC GTA AAG GAT CCA AAC CTG AAA TAC GGT        576
Val Asp Lys Ile Arg Asn Asn Val Lys Asp Pro Asn Leu Lys Tyr Gly
            180                 185                 190


GGT ACC GAC ATA GCC GTG ATA GGC CCA CCT TCT AAA GAA AAA TTC CTT        624
Gly Thr Asp Ile Ala Val Ile Gly Pro Pro Ser Lys Glu Lys Phe Leu
        195                 200                 205

AGA ATT AAT TTC CAA AGT TCC CGA GGA ACG GTC TCA CTT GGC CTA AAA        672
Arg Ile Asn Phe Gln Ser Ser Arg Gly Thr Val Ser Leu Gly Leu Lys
    210                 215                 220

CGC GAT AAC TTG TAT GTG GTC GCG TAT CTT GCA ATG GAT AAC ACG AAT        720
Arg Asp Asn Leu Tyr Val Val Ala Tyr Leu Ala Met Asp Asn Thr Asn
225                 230                 235                 240

GTT AAT CGG GCA TAT TAC TTC AAA TCA GAA ATT ACT TCC GCC GAG TTA        768
Val Asn Arg Ala Tyr Tyr Phe Lys Ser Glu Ile Thr Ser Ala Glu Leu
                245                 250                 255

ACC GCC CTT TTC CCA GAG GCC ACA ACT GCA AAT CAG AAA GCT TTA GAA        816
Thr Ala Leu Phe Pro Glu Ala Thr Thr Ala Asn Gln Lys Ala Leu Glu
            260                 265                 270

TAC ACA GAA GAT TAT CAG TCG ATC GAA AAG AAT GCC CAG ATA ACA CAG        864
Tyr Thr Glu Asp Tyr Gln Ser Ile Glu Lys Asn Ala Gln Ile Thr Gln
        275                 280                 285

GGA GAT AAA AGT AGA AAA GAA CTC GGG TTG GGG ATC GAC TTA CTT TTG        912
Gly Asp Lys Ser Arg Lys Glu Leu Gly Leu Gly Ile Asp Leu Leu Leu
    290                 295                 300

ACG TTC ATG GAA GCA GTG AAC AAG AAG GCA CGT GTG GTT AAA AAC GAA        960
Thr Phe Met Glu Ala Val Asn Lys Lys Ala Arg Val Val Lys Asn Glu
305                 310                 315                 320

GCT AGG TTT CTG CTT ATC GCT ATT CAA ATG ACA GCT GAG GTA GCA CGA       1008
Ala Arg Phe Leu Leu Ile Ala Ile Gln Met Thr Ala Glu Val Ala Arg
                325                 330                 335

TTT AGG TAC ATT CAA AAC TTG GTA ACT AAG AAC TTC CCC AAC AAG TTC       1056
Phe Arg Tyr Ile Gln Asn Leu Val Thr Lys Asn Phe Pro Asn Lys Phe
            340                 345                 350
```

```
GAC TCG GAT AAC AAG GTG ATT CAA TTT GAA GTC AGC TGG CGT AAG ATT      1104
Asp Ser Asp Asn Lys Val Ile Gln Phe Glu Val Ser Trp Arg Lys Ile
        355                 360                 365

TCT ACG GCA ATA TAC GGG GAT GCC AAA AAC GGC GTG TTT AAT AAA GAT      1152
Ser Thr Ala Ile Tyr Gly Asp Ala Lys Asn Gly Val Phe Asn Lys Asp
    370                 375                 380

TAT GAT TTC GGG TTT GGA AAA GTG AGG CAG GTG AAG GAC TTG CAA ATG      1200
Tyr Asp Phe Gly Phe Gly Lys Val Arg Gln Val Lys Asp Leu Gln Met
385                 390                 395                 400

GGA CTC CTT ATG TAT TTG GGC AAA CCA AAG                              1230
Gly Leu Leu Met Tyr Leu Gly Lys Pro Lys
                405                 410
```

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1230 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1230

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 1..465
(D) OTHER INFORMATION: /product= "bFGF"

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 472..1230
(D) OTHER INFORMATION: /product= "Saporin"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```
ATG GCT GCT GGT TCT ATC ACT ACT CTG CCG GCT CTG CCG GAA GAC GGT       48
Met Ala Ala Gly Ser Ile Thr Thr Leu Pro Ala Leu Pro Glu Asp Gly
  1               5                  10                  15

GGT TCT GGT GCT TTC CCG CCC GGC CAC TTC AAG GAC CCC AAG CGG CTG       96
Gly Ser Gly Ala Phe Pro Pro Gly His Phe Lys Asp Pro Lys Arg Leu
            20                  25                  30

TAC TGC AAA AAC GGG GGC TTC TTC CTG CGC ATC CAC CCC GAC GGC CGA      144
Tyr Cys Lys Asn Gly Gly Phe Phe Leu Arg Ile His Pro Asp Gly Arg
        35                  40                  45
```

```
GTT GAC GGG GTC CGG GAG AAG AGC GAC CCT CAC ATC AAG CTT CAA CTT        192
Val Asp Gly Val Arg Glu Lys Ser Asp Pro His Ile Lys Leu Gln Leu
     50                  55                  60

CAA GCA GAA GAG AGA GGA GTT GTG TCT ATC AAA GGA GTG TGT GCT AAC        240
Gln Ala Glu Glu Arg Gly Val Val Ser Ile Lys Gly Val Cys Ala Asn
 65                  70                  75                  80

CGT TAC CTG GCT ATG AAG GAA GAT GGA AGA TTA CTG GCT TCT AAA TGT        288
Arg Tyr Leu Ala Met Lys Glu Asp Gly Arg Leu Leu Ala Ser Lys Cys
                 85                  90                  95

GTT ACG GAT GAG TGT TTC TTT TTT GAA CGA TTG GAA TCT AAT AAC TAC        336
Val Thr Asp Glu Cys Phe Phe Phe Glu Arg Leu Glu Ser Asn Asn Tyr
            100                 105                 110

AAT ACT TAC CGG TCA AGG AAA TAC ACC AGT TGG TAT GTG GCA TTG AAA        384
Asn Thr Tyr Arg Ser Arg Lys Tyr Thr Ser Trp Tyr Val Ala Leu Lys
        115                 120                 125

CGA ACT GGG CAG TAT AAA CTT GGA TCC AAA ACA GGA CCT GGG CAG AAA        432
Arg Thr Gly Gln Tyr Lys Leu Gly Ser Lys Thr Gly Pro Gly Gln Lys
    130                 135                 140

GCT ATA CTT TTT CTT CCA ATG TCT GCT AAG AGC GCC ATG GTC ACA TCA        480
Ala Ile Leu Phe Leu Pro Met Ser Ala Lys Ser Ala Met Val Thr Ser
145                 150                 155                 160

ATC ACA TTA GAT CTA GTA AAT CCG ACC GCG GGT CAA TAC TCA TCT TTT        528
Ile Thr Leu Asp Leu Val Asn Pro Thr Ala Gly Gln Tyr Ser Ser Phe
                165                 170                 175

GTG GAT AAA ATC CGA AAC AAC GTA AAG GAT CCA AAC CTG AAA TAC GGT        576
Val Asp Lys Ile Arg Asn Asn Val Lys Asp Pro Asn Leu Lys Tyr Gly
            180                 185                 190

GGT ACC GAC ATA GCC GTG ATA GGC CCA CCT TCT AAA GAA AAA TTC CTT        624
Gly Thr Asp Ile Ala Val Ile Gly Pro Pro Ser Lys Glu Lys Phe Leu
        195                 200                 205

AGA ATT AAT TTC CAA AGT TCC CGA GGA ACG GTC TCA CTT GGC CTA AAA        672
Arg Ile Asn Phe Gln Ser Ser Arg Gly Thr Val Ser Leu Gly Leu Lys
    210                 215                 220

CGC GAT AAC TTG TAT GTG GTC GCG TAT CTT GCA ATG GAT AAC ACG AAT        720
Arg Asp Asn Leu Tyr Val Val Ala Tyr Leu Ala Met Asp Asn Thr Asn
225                 230                 235                 240

GTT AAT CGG GCA TAT TAC TTC AAA TCA GAA ATT ACT TCC GCC GAG TTA        768
Val Asn Arg Ala Tyr Tyr Phe Lys Ser Glu Ile Thr Ser Ala Glu Leu
                245                 250                 255

ACC GCC CTT TTC CCA GAG GCC ACA ACT GCA AAT CAG AAA GCT TTA GAA        816
Thr Ala Leu Phe Pro Glu Ala Thr Thr Ala Asn Gln Lys Ala Leu Glu
            260                 265                 270
```

```
TAC ACA GAA GAT TAT CAG TCG ATC GAA AAG AAT GCC CAG ATA ACA CAG       864
Tyr Thr Glu Asp Tyr Gln Ser Ile Glu Lys Asn Ala Gln Ile Thr Gln
        275                 280                 285

GGA GAT AAA AGT AGA AAA GAA CTC GGG TTG GGG ATC GAC TTA CTT TTG       912
Gly Asp Lys Ser Arg Lys Glu Leu Gly Leu Gly Ile Asp Leu Leu Leu
    290                 295                 300

ACG TTC ATG GAA GCA GTG AAC AAG AAG GCA CGT GTG GTT AAA AAC GAA       960
Thr Phe Met Glu Ala Val Asn Lys Lys Ala Arg Val Val Lys Asn Glu
305                 310                 315                 320

GCT AGG TTT CTG CTT ATC GCT ATT CAA ATG ACA GCT GAG GTA GCA CGA      1008
Ala Arg Phe Leu Leu Ile Ala Ile Gln Met Thr Ala Glu Val Ala Arg
                325                 330                 335

TTT AGG TAC ATT CAA AAC TTG GTA ACT AAG AAC TTC CCC AAC AAG TTC      1056
Phe Arg Tyr Ile Gln Asn Leu Val Thr Lys Asn Phe Pro Asn Lys Phe
            340                 345                 350

GAC TCG GAT AAC AAG GTG ATT CAA TTT GAA GTC AGC TGG CGT AAG ATT      1104
Asp Ser Asp Asn Lys Val Ile Gln Phe Glu Val Ser Trp Arg Lys Ile
        355                 360                 365

TCT ACG GCA ATA TAC GGG GAT GCC AAA AAC GGC GTG TTT AAT AAA GAT      1152
Ser Thr Ala Ile Tyr Gly Asp Ala Lys Asn Gly Val Phe Asn Lys Asp
    370                 375                 380

TAT GAT TTC GGG TTT GGA AAA GTG AGG CAG GTG AAG GAC TTG CAA ATG      1200
Tyr Asp Phe Gly Phe Gly Lys Val Arg Gln Val Lys Asp Leu Gln Met
385                 390                 395                 400

GGA CTC CTT ATG TAT TTG GGC AAA CCA AAG                              1230
Gly Leu Leu Met Tyr Leu Gly Lys Pro Lys
                405                 410
```

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 768 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 4..768
(D) OTHER INFORMATION: /product= "SAP CYS +4"

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 7..768

(D) OTHER INFORMATION: /product= "mature SAP CYS +4"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

```
CAT ATG GTC ACA TCA TGT ACA TTA GAT CTA GTA AAT CCG ACC GCG GGT      48
    Met Val Thr Ser Cys Thr Leu Asp Leu Val Asn Pro Thr Ala Gly
      1               5                  10                  15

CAA TAC TCA TCT TTT GTG GAT AAA ATC CGA AAC AAC GTA AAG GAT CCA      96
Gln Tyr Ser Ser Phe Val Asp Lys Ile Arg Asn Asn Val Lys Asp Pro
                 20                  25                  30

AAC CTG AAA TAC GGT GGT ACC GAC ATA GCC GTG ATA GGC CCA CCT TCT     144
Asn Leu Lys Tyr Gly Gly Thr Asp Ile Ala Val Ile Gly Pro Pro Ser
            35                  40                  45

AAA GAA AAA TTC CTT AGA ATT AAT TTC CAA AGT TCC CGA GGA ACG GTC     192
Lys Glu Lys Phe Leu Arg Ile Asn Phe Gln Ser Ser Arg Gly Thr Val
        50                  55                  60

TCA CTT GGC CTA AAA CGC GAT AAC TTG TAT GTG GTC GCG TAT CTT GCA     240
Ser Leu Gly Leu Lys Arg Asp Asn Leu Tyr Val Val Ala Tyr Leu Ala
    65                  70                  75

ATG GAT AAC ACG AAT GTT AAT CGG GCA TAT TAC TTC AAA TCA GAA ATT     288
Met Asp Asn Thr Asn Val Asn Arg Ala Tyr Tyr Phe Lys Ser Glu Ile
80                  85                  90                  95

ACT TCC GCC GAG TTA ACC GCC CTT TTC CCA GAG GCC ACA ACT GCA AAT     336
Thr Ser Ala Glu Leu Thr Ala Leu Phe Pro Glu Ala Thr Thr Ala Asn
                100                 105                 110

CAG AAA GCT TTA GAA TAC ACA GAA GAT TAT CAG TCG ATC GAA AAG AAT     384
Gln Lys Ala Leu Glu Tyr Thr Glu Asp Tyr Gln Ser Ile Glu Lys Asn
            115                 120                 125

GCC CAG ATA ACA CAG GGA GAT AAA AGT AGA AAA GAA CTC GGG TTG GGG     432
Ala Gln Ile Thr Gln Gly Asp Lys Ser Arg Lys Glu Leu Gly Leu Gly
        130                 135                 140

ATC GAC TTA CTT TTG ACG TTC ATG GAA GCA GTG AAC AAG AAG GCA CGT     480
Ile Asp Leu Leu Leu Thr Phe Met Glu Ala Val Asn Lys Lys Ala Arg
    145                 150                 155

GTG GTT AAA AAC GAA GCT AGG TTT CTG CTT ATC GCT ATT CAA ATG ACA     528
Val Val Lys Asn Glu Ala Arg Phe Leu Leu Ile Ala Ile Gln Met Thr
160                 165                 170                 175

GCT GAG GTA GCA CGA TTT AGG TAC ATT CAA AAC TTG GTA ACT AAG AAC     576
Ala Glu Val Ala Arg Phe Arg Tyr Ile Gln Asn Leu Val Thr Lys Asn
                180                 185                 190

TTC CCC AAC AAG TTC GAC TCG GAT AAC AAG GTG ATT CAA TTT GAA GTC     624
Phe Pro Asn Lys Phe Asp Ser Asp Asn Lys Val Ile Gln Phe Glu Val
```

```
           195                 200                 205

AGC TGG CGT AAG ATT TCT ACG GCA ATA TAC GGG GAT GCC AAA AAC GGC       672
Ser Trp Arg Lys Ile Ser Thr Ala Ile Tyr Gly Asp Ala Lys Asn Gly
        210                 215                 220

GTG TTT AAT AAA GAT TAT GAT TTC GGG TTT GGA AAA GTG AGG CAG GTG       720
Val Phe Asn Lys Asp Tyr Asp Phe Gly Phe Gly Lys Val Arg Gln Val
    225                 230                 235

AAG GAC TTG CAA ATG GGA CTC CTT ATG TAT TTG GGC AAA CCA AAG TAG       768
Lys Asp Leu Gln Met Gly Leu Leu Met Tyr Leu Gly Lys Pro Lys
240                 245                 250                 255
```

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 768 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 4..768
(D) OTHER INFORMATION: /product= "SAP CYS +10"

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 7..768
(D) OTHER INFORMATION: /product= "mature SAP CYS +10"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

```
CAT ATG GTC ACA TCA ATC ACA TTA GAT CTA GTA TGT CCG ACC GCG GGT        48
    Met Val Thr Ser Ile Thr Leu Asp Leu Val Cys Pro Thr Ala Gly
      1               5                  10                  15

CAA TAC TCA TCT TTT GTG GAT AAA ATC CGA AAC AAC GTA AAG GAT CCA        96
Gln Tyr Ser Ser Phe Val Asp Lys Ile Arg Asn Asn Val Lys Asp Pro
                20                  25                  30

AAC CTG AAA TAC GGT GGT ACC GAC ATA GCC GTG ATA GGC CCA CCT TCT       144
Asn Leu Lys Tyr Gly Gly Thr Asp Ile Ala Val Ile Gly Pro Pro Ser
            35                  40                  45

AAA GAA AAA TTC CTT AGA ATT AAT TTC CAA AGT TCC CGA GGA ACG GTC       192
Lys Glu Lys Phe Leu Arg Ile Asn Phe Gln Ser Ser Arg Gly Thr Val
        50                  55                  60

TCA CTT GGC CTA AAA CGC GAT AAC TTG TAT GTG GTC GCG TAT CTT GCA       240
```

```
Ser Leu Gly Leu Lys Arg Asp Asn Leu Tyr Val Val Ala Tyr Leu Ala
 65                  70                  75

ATG GAT AAC ACG AAT GTT AAT CGG GCA TAT TAC TTC AAA TCA GAA ATT      288
Met Asp Asn Thr Asn Val Asn Arg Ala Tyr Tyr Phe Lys Ser Glu Ile
 80                  85                  90                  95

ACT TCC GCC GAG TTA ACC GCC CTT TTC CCA GAG GCC ACA ACT GCA AAT      336
Thr Ser Ala Glu Leu Thr Ala Leu Phe Pro Glu Ala Thr Thr Ala Asn
                100                 105                 110

CAG AAA GCT TTA GAA TAC ACA GAA GAT TAT CAG TCG ATC GAA AAG AAT      384
Gln Lys Ala Leu Glu Tyr Thr Glu Asp Tyr Gln Ser Ile Glu Lys Asn
            115                 120                 125

GCC CAG ATA ACA CAG GGA GAT AAA AGT AGA AAA GAA CTC GGG TTG GGG      432
Ala Gln Ile Thr Gln Gly Asp Lys Ser Arg Lys Glu Leu Gly Leu Gly
        130                 135                 140

ATC GAC TTA CTT TTG ACG TTC ATG GAA GCA GTG AAC AAG AAG GCA CGT      480
Ile Asp Leu Leu Leu Thr Phe Met Glu Ala Val Asn Lys Lys Ala Arg
    145                 150                 155

GTG GTT AAA AAC GAA GCT AGG TTT CTG CTT ATC GCT ATT CAA ATG ACA      528
Val Val Lys Asn Glu Ala Arg Phe Leu Leu Ile Ala Ile Gln Met Thr
160                 165                 170                 175

GCT GAG GTA GCA CGA TTT AGG TAC ATT CAA AAC TTG GTA ACT AAG AAC      576
Ala Glu Val Ala Arg Phe Arg Tyr Ile Gln Asn Leu Val Thr Lys Asn
                180                 185                 190

TTC CCC AAC AAG TTC GAC TCG GAT AAC AAG GTG ATT CAA TTT GAA GTC      624
Phe Pro Asn Lys Phe Asp Ser Asp Asn Lys Val Ile Gln Phe Glu Val
            195                 200                 205

AGC TGG CGT AAG ATT TCT ACG GCA ATA TAC GGG GAT GCC AAA AAC GGC      672
Ser Trp Arg Lys Ile Ser Thr Ala Ile Tyr Gly Asp Ala Lys Asn Gly
        210                 215                 220

GTG TTT AAT AAA GAT TAT GAT TTC GGG TTT GGA AAA GTG AGG CAG GTG      720
Val Phe Asn Lys Asp Tyr Asp Phe Gly Phe Gly Lys Val Arg Gln Val
    225                 230                 235

AAG GAC TTG CAA ATG GGA CTC CTT ATG TAT TTG GGC AAA CCA AAG TAG      768
Lys Asp Leu Gln Met Gly Leu Leu Met Tyr Leu Gly Lys Pro Lys
240                 245                 250                 255
```

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..35
(A) NAME/KEY: Cathepsin B linker

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38

CCATGGCCCT GGCCCTGGCC CTGGCCCTGG CCATGG 36

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 51 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..50
(A) NAME/KEY: Cathepsin D linker

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39

CCATGGGCCG ATCGGGCTTC CTGGGCTTCG GCTTCCTGGG CTTCGCCATGG 51

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..26
(A) NAME/KEY: Gly_4Ser with NcoI ends

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40

CCATGGGCGG CGGCGGCTCT GCCATGG 27

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..41
(A) NAME/KEY: $(Gly_4Ser)_2$ with NcoI ends

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41

CCATGGGCGG CGGCGGCTCT GGCGGCGGCG GCTCTGCCAT GG 42

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 75 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..74
(A) NAME/KEY: $(Ser_4Gly)_4$ with NcoI ends

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42

CCATGGCCTC GTCGTCGTCG GGCTCGTCGT CGTCGGGCTC GTCGTCGTCG GGCTCGTCGT 60

CGTCGGGCGC CATGG 75

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..45
(A) NAME/KEY: $(Ser_4Gly)_2$

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43

CCATGGCCTC GTCGTCGTCG GGCTCGTCGT CGTCGGGCGC CATGG 45

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 96 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..95
(A) NAME/KEY: "Trypsin linker"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44

CCATGGGCCG ATCGGGCGGT GGGTGCGCTG GTAATAGAGT CAGAAGATCA GTCGGAAGCA 60

GCCTGTCTTG CGGTGGTCTC GACCTGCAGG CCATGG 96

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..18
(D) OTHER INFORMATION: /product= Thrombin substrate linker

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45

CTG GTG CCG CGC GGC AGC 18
Leu Val Pro Arg Gly Ser
1 5

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..15
(D) OTHER INFORMATION: /product= Enterokinase substrate linker

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46

GAC GAC GAC GAC CCA 15
Asp Asp Asp Asp Lys
1 5

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..12
(D) OTHER INFORMATION: /product= Factor Xa substrate

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47

ATC GAA GGT CGT 12
Ile Glu Gly Arg
1

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..8
(D) OTHER INFORMATION: /product= Flexible linker

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48

Ala Ala Pro Ala Ala Ala Pro Ala
1 5

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..4
(D) OTHER INFORMATION: /product= subtilisin substrate linker

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49

Phe Ala His Tyr
1

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..4
(D) OTHER INFORMATION: /product= subtilisin substrate linker

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50

Xaa Asp Glu Leu
1

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51

CTGGCTGCAG TTCTCTCGGC A 21

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52

TATATGCCAT GGCCAGAGTC ACTTTATCCT CCAAG 34

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53

TATATGTCGAC TATGGGAGGC TCAGCCCATGA CA 32

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54

TGAGCGAATT CCATATGGTC ACATCAATCA CATTA 35

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55

TATATGAATT CCATGGCCTT TGGTTTGCCC AAATACAT 38

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56

TATATGGATC CTATGTGTAG AGTCACTTTA TCCTCCAAG 39

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57

TATATAAGCT TCTATGGGAG GCTCAGCCCA TGACA 35

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 771 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 4..771
(D) OTHER INFORMATION: /product= "SAP CYS -1"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

```
CAT ATG TGT GTC ACA TCA ATC ACA TTA GAT CTA GTA AAT CCG ACC GCG        48
    Met Cys Val Thr Ser Ile Thr Leu Asp Leu Val Asn Pro Thr Ala
      1               5                  10                  15

GGT CAA TAC TCA TCT TTT GTG GAT AAA ATC CGA AAC AAC GTA AAA GAT        96
Gly Gln Tyr Ser Ser Phe Val Asp Lys Ile Arg Asn Asn Val Lys Asp
                 20                  25                  30

CCA AAC CTG AAA TAC GGT GGT ACC GAC ATA GCC GTG ATA GGC CCA CCT       144
Pro Asn Leu Lys Tyr Gly Gly Thr Asp Ile Ala Val Ile Gly Pro Pro
             35                  40                  45

TCT AAA GAA AAA TTC CTT AGA ATT AAT TTC CAA AGT TCC CGA GGA ACG       192
Ser Lys Glu Lys Phe Leu Arg Ile Asn Phe Gln Ser Ser Arg Gly Thr
         50                  55                  60
```

```
GTC TCA CTT GGC CTA AAA CGC GAT AAC TTG TAT GTG GTC GCG TAT CTT        240
Val Ser Leu Gly Leu Lys Arg Asp Asn Leu Tyr Val Val Ala Tyr Leu
     65                  70                  75

GCA ATG GAT AAC ACG AAT GTT AAT CGG GCA TAT TAC TTC AAA TCA GAA        288
Ala Met Asp Asn Thr Asn Val Asn Arg Ala Tyr Tyr Phe Lys Ser Glu
 80                  85                  90                  95

ATT ACT TCC GCC GAG TTA ACC GCC CTT TTC CCA GAG GCC ACA ACT GCA        336
Ile Thr Ser Ala Glu Leu Thr Ala Leu Phe Pro Glu Ala Thr Thr Ala
                100                 105                 110

AAT CAG AAA GCT TTA GAA TAC ACA GAA GAT TAT CAG TCG ATC GAA AAG        384
Asn Gln Lys Ala Leu Glu Tyr Thr Glu Asp Tyr Gln Ser Ile Glu Lys
            115                 120                 125

AAT GCC CAG ATA ACA CAG GGA GAT AAA AGT AGA AAA GAA CTC GGG TTG        432
Asn Ala Gln Ile Thr Gln Gly Asp Lys Ser Arg Lys Glu Leu Gly Leu
        130                 135                 140

GGG ATC GAC TTA CTT TTG ACG TTC ATG GAA GCA GTG AAC AAG AAG GCA        480
Gly Ile Asp Leu Leu Leu Thr Phe Met Glu Ala Val Asn Lys Lys Ala
    145                 150                 155

CGT GTG GTT AAA AAC GAA GCT AGG TTT CTG CTT ATC GCT ATT CAA ATG        528
Arg Val Val Lys Asn Glu Ala Arg Phe Leu Leu Ile Ala Ile Gln Met
160                 165                 170                 175

ACA GCT GAG GTA GCA CGA TTT AGG TAC ATT CAA AAC TTG GTA ACT AAG        576
Thr Ala Glu Val Ala Arg Phe Arg Tyr Ile Gln Asn Leu Val Thr Lys
                180                 185                 190

AAC TTC CCC AAC AAG TTC GAC TCG GAT AAC AAG GTG ATT CAA TTT GAA        624
Asn Phe Pro Asn Lys Phe Asp Ser Asp Asn Lys Val Ile Gln Phe Glu
            195                 200                 205

GTC AGC TGG CGT AAG ATT TCT ACG GCA ATA TAC GGG GAT GCC AAA AAC        672
Val Ser Trp Arg Lys Ile Ser Thr Ala Ile Tyr Gly Asp Ala Lys Asn
        210                 215                 220

GGC GTG TTT AAT AAA GAT TAT GAT TTC GGG TTT GGA AAA GTG AGG CAG        720
Gly Val Phe Asn Lys Asp Tyr Asp Phe Gly Phe Gly Lys Val Arg Gln
    225                 230                 235

GTG AAG GAC TTG CAA ATG GGA CTC CTT ATG TAT TTG GGC AAA CCA AAG        768
Val Lys Asp Leu Gln Met Gly Leu Leu Met Tyr Leu Gly Lys Pro Lys
240                 245                 250                 255

TAG                                                                    771
```

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

CATATGGTCA CATCATGTAC ATTAGATCTA GTAAAT 36

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

CATATGGTCA CATCAATCAC ATTAGATCTA GTATGTCCGA CCGCGGGTCA 50

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

TTTCAGGTTT GGATCTTTTA CGTTGTTT 28

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

AAACAACGTA AAAGATCCAA ACCTGAAA 28

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..7
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63

Ala Pro Arg Arg Arg Lys Leu
1 5

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

Lys Arg Lys Lys Lys
1 5

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..5
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65

Ile Arg Val Arg Arg
1 5

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..6
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66

Lys Arg Lys Arg Lys Lys
1 5

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..7
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67

Pro Lys Lys Arg Lys Val Glu
1 5

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..8
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68

Pro Pro Lys Lys Ala Arg Glu Val
1 5

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..9
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69

Pro Ala Ala Lys Arg Val Lys Leu Asp
1 5

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..5
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70

Lys Arg Pro Arg Pro
1 5

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS

(B) LOCATION: 1..5
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71

Lys Ile Pro Ile Lys
1 5

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..9
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72

Gly Lys Arg Lys Arg Lys Ser
1 5

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..9
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73

Ser Lys Arg Val Ala Lys Arg Lys leu
1 5

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..9
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74

Ser His Trp Lys Gln Lys Arg Lys Phe
1 5

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..8
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75

Pro Leu Leu Lys Lys Ile Lys Gln
1 5

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..7
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76

Pro Gln Pro Lys Lys Lys Pro
1 5

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..15
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77

Pro Gly Lys Arg Lys Lys Glu Met Thr Lys Gln Lys Glu Val Pro
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..12
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78

Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg Ala Pro
1 5 10

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..7
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79

Asn Tyr Lys Lys Pro Lys Leu
1 5

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..7
(D) OTHER INFORMATION: /product= nuclear translocation sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80

His Phe Lys Asp Pro Lys Arg
1 5

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 783 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 10..783
(D) OTHER INFORMATION: /product= "Amplified SAP with EcoR1 ends"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

```
GAATTCCAT ATG GTC ACA TCA ATC ACA TTA GAT CTA GTA AAT CCG ACC          48
          Met Val Thr Ser Ile Thr Leu Asp Leu Val Asn Pro Thr
            1               5                  10

GCG GGT CAA TAC TCA TCT TTT GTG GAT AAA ATC CGA AAC AAC GTA AAG        96
Ala Gly Gln Tyr Ser Ser Phe Val Asp Lys Ile Arg Asn Asn Val Lys
     15                  20                  25

GAT CCA AAC CTG AAA TAC GGT GGT ACC GAC ATA GCC GTG ATA GGC CCA       144
Asp Pro Asn Leu Lys Tyr Gly Gly Thr Asp Ile Ala Val Ile Gly Pro
 30                  35                  40                  45
```

```
CCT TCT AAA GAA AAA TTC CTT AGA ATT AAT TTC CAA AGT TCC CGA GGA        192
Pro Ser Lys Glu Lys Phe Leu Arg Ile Asn Phe Gln Ser Ser Arg Gly
                  50                  55                  60

ACG GTC TCA CTT GGC CTA AAA CGC GAT AAC TTG TAT GTG GTC GCG TAT        240
Thr Val Ser Leu Gly Leu Lys Arg Asp Asn Leu Tyr Val Val Ala Tyr
              65                  70                  75

CTT GCA ATG GAT AAC ACG AAT GTT AAT CGG GCA TAT TAC TTC AAA TCA        288
Leu Ala Met Asp Asn Thr Asn Val Asn Arg Ala Tyr Tyr Phe Lys Ser
          80                  85                  90

GAA ATT ACT TCC GCC GAG TTA ACC GCC CTT TTC CCA GAG GCC ACA ACT        336
Glu Ile Thr Ser Ala Glu Leu Thr Ala Leu Phe Pro Glu Ala Thr Thr
      95                  100                 105

GCA AAT CAG AAA GCT TTA GAA TAC ACA GAA GAT TAT CAG TCG ATC GAA        384
Ala Asn Gln Lys Ala Leu Glu Tyr Thr Glu Asp Tyr Gln Ser Ile Glu
110                 115                 120                 125

AAG AAT GCC CAG ATA ACA CAG GGA GAT AAA AGT AGA AAA GAA CTC GGG        432
Lys Asn Ala Gln Ile Thr Gln Gly Asp Lys Ser Arg Lys Glu Leu Gly
                  130                 135                 140

TTG GGG ATC GAC TTA CTT TTG ACG TTC ATG GAA GCA GTG AAC AAG AAG        480
Leu Gly Ile Asp Leu Leu Leu Thr Phe Met Glu Ala Val Asn Lys Lys
              145                 150                 155

GCA CGT GTG GTT AAA AAC GAA GCT AGG TTT CTG CTT ATC GCT ATT CAA        528
Ala Arg Val Val Lys Asn Glu Ala Arg Phe Leu Leu Ile Ala Ile Gln
          160                 165                 170

ATG ACA GCT GAG GTA GCA CGA TTT AGG TAC ATT CAA AAC TTG GTA ACT        576
Met Thr Ala Glu Val Ala Arg Phe Arg Tyr Ile Gln Asn Leu Val Thr
      175                 180                 185

AAG AAC TTC CCC AAC AAG TTC GAC TCG GAT AAC AAG GTG ATT CAA TTT        624
Lys Asn Phe Pro Asn Lys Phe Asp Ser Asp Asn Lys Val Ile Gln Phe
190                 195                 200                 205

GAA GTC AGC TGG CGT AAG ATT TCT ACG GCA ATA TAC GGG GAT GCC AAA        672
Glu Val Ser Trp Arg Lys Ile Ser Thr Ala Ile Tyr Gly Asp Ala Lys
                  210                 215                 220

AAC GGC GTG TTT AAT AAA GAT TAT GAT TTC GGG TTT GGA AAA GTG AGG        720
Asn Gly Val Phe Asn Lys Asp Tyr Asp Phe Gly Phe Gly Lys Val Arg
              225                 230                 235

CAG GTG AAG GAC TTG CAA ATG GGA CTC CTT ATG TAT TTG GGC AAA CCA        768
Gln Val Lys Asp Leu Gln Met Gly Leu Leu Met Tyr Leu Gly Lys Pro
          240                 245                 250

AAG GCC ATG GAA TTC                                                    783
Lys Ala Met Glu Phe
      255
```

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1005 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1005
(D) OTHER INFORMATION: /product= "SAP-HBEGF"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

```
ATG GTC ACA TCA ATC ACA TTA GAT CTA GTA AAT CCG ACC GCG GGT CAA        48
Met Val Thr Ser Ile Thr Leu Asp Leu Val Asn Pro Thr Ala Gly Gln
  1               5                  10                  15

TAC TCA TCT TTT GTG GAT AAA ATC CGA AAC AAC GTA AAG GAT CCA AAC        96
Tyr Ser Ser Phe Val Asp Lys Ile Arg Asn Asn Val Lys Asp Pro Asn
             20                  25                  30

CTG AAA TAC GGT GGT ACC GAC ATA GCC GTG ATA GGC CCA CCT TCT AAA       144
Leu Lys Tyr Gly Gly Thr Asp Ile Ala Val Ile Gly Pro Pro Ser Lys
         35                  40                  45

GAA AAA TTC CTT AGA ATT AAT TTC CAA AGT TCC CGA GGA ACG GTC TCA       192
Glu Lys Phe Leu Arg Ile Asn Phe Gln Ser Ser Arg Gly Thr Val Ser
     50                  55                  60

CTT GGC CTA AAA CGC GAT AAC TTG TAT GTG GTC GCG TAT CTT GCA ATG       240
Leu Gly Leu Lys Arg Asp Asn Leu Tyr Val Val Ala Tyr Leu Ala Met
 65                  70                  75                  80

GAT AAC ACG AAT GTT AAT CGG GCA TAT TAC TTC AAA TCA GAA ATT ACT       288
Asp Asn Thr Asn Val Asn Arg Ala Tyr Tyr Phe Lys Ser Glu Ile Thr
                 85                  90                  95

TCC GCC GAG TTA ACC GCC CTT TTC CCA GAG GCC ACA ACT GCA AAT CAG       336
Ser Ala Glu Leu Thr Ala Leu Phe Pro Glu Ala Thr Thr Ala Asn Gln
            100                 105                 110

AAA GCT TTA GAA TAC ACA GAA GAT TAT CAG TCG ATC GAA AAG AAT GCC       384
Lys Ala Leu Glu Tyr Thr Glu Asp Tyr Gln Ser Ile Glu Lys Asn Ala
        115                 120                 125

CAG ATA ACA CAG GGA GAT AAA AGT AGA AAA GAA CTC GGG TTG GGG ATC       432
Gln Ile Thr Gln Gly Asp Lys Ser Arg Lys Glu Leu Gly Leu Gly Ile
    130                 135                 140
```

```
GAC TTA CTT TTG ACG TTC ATG GAA GCA GTG AAC AAG AAG GCA CGT GTG        480
Asp Leu Leu Leu Thr Phe Met Glu Ala Val Asn Lys Lys Ala Arg Val
145                 150                 155                 160

GTT AAA AAC GAA GCT AGG TTT CTG CTT ATC GCT ATT CAA ATG ACA GCT        528
Val Lys Asn Glu Ala Arg Phe Leu Leu Ile Ala Ile Gln Met Thr Ala
                165                 170                 175

GAG GTA GCA CGA TTT AGG TAC ATT CAA AAC TTG GTA ACT AAG AAC TTC        576
Glu Val Ala Arg Phe Arg Tyr Ile Gln Asn Leu Val Thr Lys Asn Phe
            180                 185                 190

CCC AAC AAG TTC GAC TCG GAT AAC AAG GTG ATT CAA TTT GAA GTC AGC        624
Pro Asn Lys Phe Asp Ser Asp Asn Lys Val Ile Gln Phe Glu Val Ser
        195                 200                 205

TGG CGT AAG ATT TCT ACG GCA ATA TAC GGG GAT GCC AAA AAC GGC GTG        672
Trp Arg Lys Ile Ser Thr Ala Ile Tyr Gly Asp Ala Lys Asn Gly Val
    210                 215                 220

TTT AAT AAA GAT TAT GAT TTC GGG TTT GGA AAA GTG AGG CAG GTG AAG        720
Phe Asn Lys Asp Tyr Asp Phe Gly Phe Gly Lys Val Arg Gln Val Lys
225                 230                 235                 240

GAC TTG CAA ATG GGA CTC CTT ATG TAT TTG GGC AAA CCA AAG GCC ATG        768
Asp Leu Gln Met Gly Leu Leu Met Tyr Leu Gly Lys Pro Lys Ala Met
                245                 250                 255

GCC AGA GTC ACT TTA TCC TCC AAG CCA CAA GCA CTG GCC ACA CCA AAC        816
Ala Arg Val Thr Leu Ser Ser Lys Pro Gln Ala Leu Ala Thr Pro Asn
            260                 265                 270

AAG GAG GAG CAC GGG AAA AGA AAG AAG AAA GGC AAG GGG CTA GGG AAG        864
Lys Glu Glu His Gly Lys Arg Lys Lys Lys Gly Lys Gly Leu Gly Lys
        275                 280                 285

AAG AGG GAC CCA TGT CTT CGG AAA TAC AAG GAC TTC TGC ATC CAC GGA        912
Lys Arg Asp Pro Cys Leu Arg Lys Tyr Lys Asp Phe Cys Ile His Gly
    290                 295                 300

GAA TGC AAA TAT GTG AAG GAG CTC CGG GCT CCC TCC TGC ATC TGC CAC        960
Glu Cys Lys Tyr Val Lys Glu Leu Arg Ala Pro Ser Cys Ile Cys His
305                 310                 315                 320

CCG GGT TAT CAT GGA GAG AGG TGT CAT GGG CTG AGC CTC CCA TA             1005
Pro Gly Tyr His Gly Glu Arg Cys His Gly Leu Ser Leu Pro
                325                 330                 335
```

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 240 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..240
(D) OTHER INFORMATION: /product= "MET-CYS-HBEGF"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

```
ATG TGT AGA GTC ACT TTA TCC TCC AAG CCA CAA GCA CTG GCC ACA CCA        48
Met Cys Arg Val Thr Leu Ser Ser Lys Pro Gln Ala Leu Ala Thr Pro
  1               5                  10                  15

AAC AAG GAG GAG CAC GGG AAA AGA AAG AAG AAA GGC AAG GGG CTA GGG        96
Asn Lys Glu Glu His Gly Lys Arg Lys Lys Lys Gly Lys Gly Leu Gly
             20                  25                  30

AAG AAG AGG GAC CCA TGT CTT CGG AAA TAC AAG GAC TTC TGC ATC CAC       144
Lys Lys Arg Asp Pro Cys Leu Arg Lys Tyr Lys Asp Phe Cys Ile His
         35                  40                  45

GGA GAA TGC AAA TAT GTG AAG GAG CTC CGG GCT CCC TCC TGC ATC TGC       192
Gly Glu Cys Lys Tyr Val Lys Glu Leu Arg Ala Pro Ser Cys Ile Cys
     50                  55                  60

CAC CCG GGT TAT CAT GGA GAG AGG TGT CAT GGG CTG AGC CTC CCA TAG       240
His Pro Gly Tyr His Gly Glu Arg Cys His Gly Leu Ser Leu Pro
 65                  70                  75                  80
```

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 249 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..249
(D) OTHER INFORMATION: /product=
"MET-CYS-ALA-MET-ALA-HBEGF"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

```
ATG TGT GCC ATG GCC AGA GTC ACT TTA TCC TCC AAG CCA CAA GCA CTG        48
Met Cys Ala Met Ala Arg Val Thr Leu Ser Ser Lys Pro Gln Ala Leu
  1               5                  10                  15

GCC ACA CCA AAC AAG GAG GAG CAC GGG AAA AGA AAG AAG AAA GGC AAG        96
Ala Thr Pro Asn Lys Glu Glu His Gly Lys Arg Lys Lys Lys Gly Lys
             20                  25                  30

GGG CTA GGG AAG AAG AGG GAC CCA TGT CTT CGG AAA TAC AAG GAC TTC       144
Gly Leu Gly Lys Lys Arg Asp Pro Cys Leu Arg Lys Tyr Lys Asp Phe
         35                  40                  45

TGC ATC CAC GGA GAA TGC AAA TAT GTG AAG GAG CTC CGG GCT CCC TCC       192
Cys Ile His Gly Glu Cys Lys Tyr Val Lys Glu Leu Arg Ala Pro Ser
     50                  55                  60

TGC ATC TGC CAC CCG GGT TAT CAT GGA GAG AGG TGT CAT GGG CTG AGC       240
Cys Ile Cys His Pro Gly Tyr His Gly Glu Arg Cys His Gly Leu Ser
 65                  70                  75                  80

CTC CCA TAG                                                           249
Leu Pro
```

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

```
Met Cys Ala Met Ala
  1               5
```

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 51 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

```
TATAGGATCC TGATGTGTGC CATGGCCAGA GTCACTTTAT CCTCCAAGCC A              51
```

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

Ile Lys Arg Leu Arg Arg
1 5

Claims

1. A conjugate, comprising a targeted agent and a heparin-binding epidermal-like growth factor (HBEGF) polypeptide or a portion thereof, wherein the conjugate binds to a HBEGF receptor resulting in internalization of the linked targeted agent in cells bearing the receptor.

2. A conjugate, comprising the following components: $(HBEGF)_n$, $(L)_q$ and $(\text{targeted agent})_m$, wherein:

L is a linker;

HBEGF is a HBEGF polypeptide;

at least one HBEGF polypeptide is linked at any residue in the polypeptide via $(L)_q$ to at least one targeted agent;

m and n, which are selected independently, are at least 1;

q is 0 or more as long as the resulting conjugate binds to the targeted receptor, is internalized and delivers the targeted agent; and

the conjugate binds to a receptor that interacts with and internalizes HBEGF, whereby the targeted agent(s) is internalized in a cell bearing the receptor.

3. The conjugate of claim 2, wherein m and n, which are selected independently, are from 1 to 6.

4. The conjugate of claim 2, wherein q is 1, n is 1 and m is 1.

5. The conjugate of claim 2, wherein L is selected from the group consisting of protease substrates, linkers that increase the flexibility of the conjugate, linkers that increase the solubility of the conjugate, linkers that increase the serum stability of the conjugate, photocleavable linkers and acid cleavable linkers.

6. The conjugate of claim 5, wherein the linker is selected from the group consisting of cathepsin B substrate, cathepsin D substrate, trypsin substrate, thrombin substrate, recombinant subtilisin substrate, $(Gly_mSer_p)_n$, $(Ser_mGly_p)_n$ and $(AlaAlaProAla)_n$ in which n is 1 to 6, m is 1 to 6 and p is 1 to 4.

7. The conjugate of claim 2, wherein L is selected from the group consisting of $(Gly_4Ser)_q$ and $(Ser_4Gly)_q$, where q is 1 to 4.

8. A conjugate of any one of claim 1 or 2, wherein a HBEGF polypeptide is selected from the group consisting of mammalian HBEGF polypeptides and HBEGF polypeptides in which a cysteine residue is added or replaces a non-essential amino acid residue within about 20 amino acids of the N-terminus or C-terminus of the polypeptide.

9. The conjugate of any one of claim 1 or 2, wherein the targeted agent is a cytotoxic agent.

10. The conjugate of any one of claim 1 or 2, wherein the targeted agent is a ribosome-inactivating protein.

11. The conjugate of claim 10, wherein the targeted agent is a saporin.

12. The conjugate of any one of claim 1 or 2, wherein the targeted agent is a nucleic acid.

13. The conjugate of claim 2, wherein the conjugate that is a fusion protein selected from the group consisting of FPH1, FPHS1, FPHS2,FPHS3, FPHS4, FPHS5, FPSH1 and FPSH2.

14. A conjugate comprising a polypeptide of the formula: targeted $agent_n$-$(L)_q$-$HBEGF_m$ or $HBEGF_m$-$(L)_q$-targeted $agent_n$, wherein:

L is a linker;

HBEGF is a HBEGF polypeptide;

at least one HBEGF polypeptide is linked at any residue in the polypeptide via $(L)_q$ to at least one targeted agent;

m and n, which are selected independently, are at least 1;

q is 0 or more as long as the resulting conjugate binds to the targeted receptor. is internalized and delivers the targeted agent; and

the conjugate binds to a receptor that interacts with and internalizes HBEGF, whereby the targeted agent(s) is internalized in a cell bearing the receptor.

15. The conjugate of claim 14, wherein each HBEGF polypeptide in the conjugate is independently selected from the group consisting of HBEGF polypeptides and HBEGF polypeptides in which a cysteine residue is added or replaces a non-essential amino acid residue within about 20 amino acids of the N-terminus of the polypeptide.

16. The conjugate of any one of claims 1-15 for use as an active therapeutic substance.

17. The conjugate of any one of claims 1-15 for use in the manufacture of a medicament for treating an HBEGF-mediated pathophysiological condition.

18. The conjugate of claim 17, wherein the pathophysiological condition is a solid tumor in which the cells bear receptors to which HBEGF binds, a dermatological disorder involving epidermal cells, an ophthalmic disorder involving proliferation of epithelial cells, or a disorder characterized by proliferation of smooth muscle cells.

19. The conjugate of any one of claims 1-15 for use in inhibiting proliferation of cells bearing HBEGF receptors.

20. The conjugate of claim 12, for use in effecting gene therapy, wherein the conjugate includes a nuclear translocation sequence operatively linked to the targeted nucleic acid or to a HBEGF.

21. A DNA fragment comprising a sequence of nucleotides encoding the conjugate of any one of claims 1-4 and 6-15.

22. A plasmid, comprising the DNA of claim 21.

23. The plasmid of claim 22, wherein the plasmid is an expression vector for expression of the DNA encoding the conjugate in eukaryotic cells or is an expression vector for expression of the conjugate in prokaryotic cells.

24. The expression vector of claim 23, comprising a DNA encoding a secretion signal sequence operatively linked to the DNA encoding the conjugate.

25. The expression vector of claim 24, wherein the secretion signal is selected from the group consisting of OmpA, OmpT, phoA, bacterial alkaline phosphatase, pelB, the insulin leader sequence, mammalian alkaline phosphatase, growth hormone leader sequence and mellitin.

26. The plasmid of claim 22 that is selected from the group consisting of PZ30B1, PZ31B1, PZ32B1, PZ33B1, PZ34B1, PZ35B1, PZ36B1 and PZ37B1.

27. A cell transfected or transformed with the expression vector of claim 23.

28. The cell of claim 27 that is a bacterial cell.

29. The cell of claim 27 that is an insect cell.

30. A method of producing a HBEGF conjugate, comprising culturing the cells of claim 27 under conditions whereby DNA is transcribed and translated to produce the conjugate.

31. A heparin-binding epidermal growth factor-like growth factor (HBEGF) polypeptide that is modified by insertion of a cysteine residue or methionine-cysteine within or at about twenty amino acids of the N-terminus or C-terminus, wherein the inserted residue replaces a nonessential residue in an unmodified HBEGF polypeptide or is added to the HBEGF polypeptide.

32. The modified polypeptide of claim 31, wherein the cysteine residue is inserted within or at about 10 residues from the N-terminus.

33. The modified polypeptide of claim 32 that is Met-Cys-HBEGF.

34. A DNA fragment comprising a sequence of nucleotides encoding the modified polypeptide of claim 31.

35. A plasmid selected from the group consisting of PZ38I, PZ39I, PZ40I and PZ41I.

36. The modified polypeptide of claim 31, selected from the group consisting of FPSH2, FPSH3, FPSH4 and FPSH5.

37. A pharmaceutical composition comprising a conjugate according to any one of claims 1-4 or 6-16, in combination with a physiologically acceptable excipient.

38. A method of producing an HBEGF fusion protein, comprising:

(a) culturing cells transformed with a plasmid containing a DNA fragment according to claim 21, under conditions whereby the DNA fragment is transcribed and translated;

(b) lysing the transformed cells in a buffer containing urea to form a lysate containing an HBEGF fusion protein;

(c) applying the lysate to a cation-exchange chromatography resin;

(d) eluting the HBEGF fusion protein from the cation-exchange chromatography resin of step (c);

(e) passing the HBEGF fusion protein over an anion-exchange chromatography resin;

(f) applying the HBEGF fusion protein to an anion-exchange chromatography resin;

(g) eluting the HBEGF fusion protein from the cation-exchange chromatography resin of step (f);

(h) applying the HBEGF fusion protein to a hydrophobic interaction chromatography resin;

(i) eluting the HBEGF fusion protein from the hydrophobic interaction chromatography resin of step (h); and

(j) recovering the HBEGF fusion protein from a size exclusion chromatography resin.